Deep Learning: Why and What Case Study: Image Classification

Hao Zhang haomoodzhang@gmail.com

LAMDA Group
National Key Laboratory for Novel Software Technology
Department of Computer Science and Technology
Nanjing University

May 6, 2016

Table of Contents I



Challenges in Image Recognition and Data-Driven Approach

Image Classification

Challenges in Image Recognition

Semantic Gap
Other Challenges

Data-Driven Approach

Linear Classifier

Training Data

Hypothesis Function

Hypothesis of Linear Classifier

Understanding of Hypothesis Function Hypothesis of Softmax Classifier

Loss Function

Cross Entropy Loss

Regularization

Optimization

Gradient Descent

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function

Cross Entropy Loss

Regularization

Gradient Descent Softmax Update

1) Rule Prediction and

Table of Contents II



Softmax Update Rule
Prediction and Evaluation
Prediction
Evaluation

Feedforward Neural Networks

Representation Learning

Linear Models

Feature Engineering

Kernel Methods

Representation Learning

Deep Learning

Hypothesis Function

Data Representation
Artificial Neural Networks

Architecture

Loss Function Optimization

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function

Cross Entropy Loss

Regularization

Gradient Descent Softmax Update

Rule

Table of Contents III



Gradient Descent
Error Back-Propagation
Prediction and Evaluation

Convolutional Neural Networks

Training Data
Hypothesis Function
Convolutional Layer
Pooling Layer
ReLU Layer
Fully Connected Layer

References

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function

Cross Entropy Loss

Regularization

Prediction and

Gradient Descens
Softmax Update

Table of Contents



Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Data-Driven Approach

Linear Classifier

Feedforward Neural Networks

Convolutional Neural Networks

References

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss

Regularization

Softmax Update
Rule
Prediction and

Image Classification



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition

Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven

inear Classifier

Approach

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier

Cross Entropy Loss Regularization Optimization Gradient Descent

Loss Function

Softmax Update
Rule
Prediction and

Image Classification



Given a fixed categories, such as {cat, dog, plane, trunk, ...},

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai
Challenges in

Image Recognition and Data-Driven Approach

Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

Prediction and

Softmax Update Rule

Image Classification



- ▶ Given a fixed categories, such as {cat, dog, plane, trunk, ...},
- ► For each image, assign one of those categories as its label.

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Recognition and Data-Driven Approach

6) Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

Prediction and

Softmax Update Rule



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition

Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of

Hypothesis of Softmax Classifier Loss Function Cross Entropy Loss Regularization

Optimization Gradient Descen-Softmax Update

1 Rule

Image recognition task



Image recognition task

Easy for human,

Why and What Hao Zhang haomoodzhang@gmai

Deep Learning:

Image Classification Challenges in Image Recognition

Semantic Gap Approach

Training Data Hypothesis of Linear Understanding of Hypothesis of Softmax Classifier

Loss Function Cross Entropy Loss Regularization Gradient Descent

Softmax Update Rule

Prediction and



Image recognition task

- Easy for human,
- Hard for computer,

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition

Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function

Cross Entropy Loss

Regularization
Optimization
Gradient Descent
Softmax Update

Rule Prediction and



Image recognition task

- Easy for human,
- Hard for computer,
- Computer can only see an array of numbers.

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition

Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier

Gradient Descent Softmax Update Rule

Prediction and

Loss Function Cross Entropy Loss Regularization

Other Challenges



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai



Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap

Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function

Hypothesis of Softmax Classifier Loss Function Cross Entropy Loss

Regularization
Optimization
Gradient Descent
Softmax Update

71 Rule

Figure: Other challenges for image recognition. This figure is reproduced from [3].

How to do image classification?



Why and What Hao Zhang haomoodzhang@gmai

Deep Learning:

Image Classification Challenges in Image Recognition Semantic Gap

Data-Driven

Approach

Training Data

Hypothesis of Linear Understanding of

Hypothesis of Softmax Classifier Loss Function Cross Entropy Loss Regularization

Gradient Descent

Prediction and

Softmax Update Rule



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap

Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function

Softmax Classifier Loss Function Cross Entropy Loss Regularization

Optimization

Gradient Descent

Softmax Update

Rule Prediction and

How to do image classification?

Can not be solved by rule-based methods,



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Image Classification Challenges in Image Semantic Gap

Data-Driven

Approach

Training Data Hypothesis of Linear Understanding of

Loss Function Cross Entropy Loss

Regularization

Rule

Softmax Update

Prediction and

How to do image classification?

- Can not be solved by rule-based methods,
- So we mimic the human learning process,



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Image Classification Challenges in Image Semantic Gap

Data-Driven

Approach

Hypothesis Function Hypothesis of Linear Understanding of

Loss Function Cross Entropy Loss

Regularization

Softmax Update Rule Prediction and

How to do image classification?

- Can not be solved by rule-based methods,
- So we mimic the human learning process,
- Let machine learn from data.

Table of Contents



Challenges in Image Recognition and Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Loss Function
Optimization
Prediction and Evaluation

Feedforward Neural Networks

Convolutional Neural Networks

References

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

10 Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss

Gradient Descens

Regularization

Rule



$$D = \{ (\vec{x}^{(i)}, y^{(i)}) \}_{i=1}^{m},$$
 (1)

$$\vec{x}^{(i)} \in \mathbb{R}^n, \forall i,$$
 (2)

$$y^{(i)} \in \{0, 1, 2, 3, \dots, K - 1\}, \forall i.$$
 (3)

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai
Challenges in

Challenges in
Image Recognition
and Data-Driven
Approach
Image Classification
Challenges in Image

Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss

Regularization
Optimization
Gradient Descent
Softmax Update
Rule
Prediction and



$$D = \{ (\vec{x}^{(i)}, y^{(i)}) \}_{i=1}^{m}, \qquad (1)$$

$$\vec{x}^{(i)} \in \mathbb{R}^n, \forall i,$$
 (2)

$$y^{(i)} \in \{0, 1, 2, 3, \dots, K - 1\}, \forall i.$$
 (3)

▶ m: training data size,

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Approach
Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss

Softmax Update
Rule
Prediction and

Regularization



$$D = \{ (\vec{x}^{(i)}, y^{(i)}) \}_{i=1}^{m}, \qquad (1)$$

$$\vec{x}^{(i)} \in \mathbb{R}^n, \forall i,$$
 (2)

$$y^{(i)} \in \{0, 1, 2, 3, \dots, K - 1\}, \forall i.$$
 (3)

- m: training data size,
- K: total classes,

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Classification Challenges in Image

Recognition Semantic Gap Approach

Training Data Hypothesis of Linear Understanding of Hypothesis of Loss Function Cross Entropy Loss

Softmax Update Rule

Prediction and

Regularization



$$D = \{ (\vec{x}^{(i)}, y^{(i)}) \}_{i=1}^{m}, \qquad (1)$$

$$\vec{x}^{(i)} \in \mathbb{R}^n, \forall i,$$
 (2)

$$y^{(i)} \in \{0, 1, 2, 3, \dots, K - 1\}, \forall i.$$
 (3)

- ▶ m: training data size,
- K: total classes,
- $\vec{x}^{(i)}$: *i*-th input image, expand it into vector,

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai Challenges in

Image Recognition and Data-Driven Approach Image Classification Challenges in Image

Data-Driven Approach

Semantic Gap

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization
Ontimization

Softmax Update
Rule
Prediction and



$$D = \{ (\vec{x}^{(i)}, y^{(i)}) \}_{i=1}^{m}, \qquad (1)$$

$$\vec{x}^{(i)} \in \mathbb{R}^n, \forall i,$$
 (2)

$$y^{(i)} \in \{0, 1, 2, 3, \dots, K - 1\}, \forall i.$$
 (3)

- m: training data size,
- K: total classes.
- $ightharpoonup \vec{x}^{(i)}$: *i*-th input image, expand it into vector,
- ▶ n: dimension of input data. E.g., if input image size $1024 \times 768 \times 3$, then $n = 1024 \times 768 \times 32 = 359296$.

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Classification

Semantic Gap Approach

Challenges in Image

Training Data Hypothesis of Linear Understanding of Loss Function Cross Entropy Loss Regularization

Softmax Update Rule

Prediction and



$$D = \{ (\vec{x}^{(i)}, y^{(i)}) \}_{i=1}^{m},$$
 (1)

$$\vec{x}^{(i)} \in \mathbb{R}^n, \forall i,$$
 (2)

$$y^{(i)} \in \{0, 1, 2, 3, \dots, K - 1\}, \forall i.$$
 (3)

- ▶ m: training data size,
- K: total classes,
- $ightharpoonup \vec{x}^{(i)}$: *i*-th input image, expand it into vector,
- ▶ n: dimension of input data. E.g., if input image size $1024 \times 768 \times 3$, then $n = 1024 \times 768 \times 32 = 359296$,
- $y^{(i)}$: label corresponds to $\vec{x}^{(i)}$.

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach Image Classification

Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

IL) Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

Softmax Update Rule

Hypothesis Function



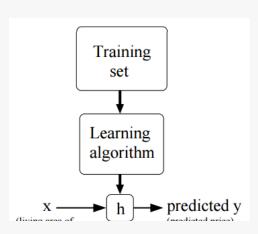


Figure: $h(\vec{x})$ gives confidence or score of \vec{x} belonging to different classes. This figure is reproduced from [8].

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven

Linear Classifier

Approach

Training Data

Hypothesis Function
Hypothesis of Linear
Classifier

Understanding of Hypothesis Function Hypothesis of Softmax Classifier Loss Function Cross Entropy Loss

Regularization
Optimization
Gradient Descent
Softmax Update

Rule Prediction and

h: Linear mapping



Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Classification Challenges in Image Recognition Semantic Gap Data-Driven Approach

Training Data Hypothesis Function Hypothesis of Linear Classifier

Understanding of

Hypothesis of Softmax Classifier Loss Function Cross Entropy Loss Regularization

Gradient Descent Softmax Update Rule

Prediction and



h: Linear mapping

$$h(\vec{x}; W, \vec{b}) = W\vec{x} + \vec{b}, \qquad (4)$$

$$W \in \mathbb{R}^{K \times n}, \vec{b} \in \mathbb{R}^{K}$$
.

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Image Classification Challenges in Image Semantic Gap

(5)

Data-Driven

Approach

Training Data Hypothesis Function Hypothesis of Linear Classifier

Understanding of

Softmax Classifier Loss Function Cross Entropy Loss Regularization

Gradient Descent

Softmax Update Rule Prediction and



h: Linear mapping

$$h(\vec{x}; W, \vec{b}) = W\vec{x} + \vec{b}, \qquad (4)$$

$$W \in \mathbb{R}^{K \times n}, \vec{b} \in \mathbb{R}^{K}$$
.

 \blacktriangleright (W, \vec{b}) : parameters of h,

Why and What Hao Zhang haomoodzhang@gmai

Deep Learning:

Image Classification Challenges in Image Semantic Gap Approach

(5)

Training Data

Hypothesis of Linear Classifier Understanding of

Hypothesis of Softmax Classifier Loss Function Cross Entropy Loss Regularization

Softmax Update Rule

Prediction and



h: Linear mapping

$$h(\vec{x}; W, \vec{b}) = W\vec{x} + \vec{b}, \qquad (4)$$

$$W \in \mathbb{R}^{K \times n}, \vec{b} \in \mathbb{R}^{K}$$
.

- \blacktriangleright (W, \vec{b}) : parameters of h,
- ▶ W: weights,

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Image Classification Challenges in Image Semantic Gap Approach

(5)

Training Data Hypothesis of Linear Classifier

Understanding of Hypothesis of Loss Function Cross Entropy Loss Regularization

Softmax Update Rule Prediction and



h: Linear mapping

$$h(\vec{x}; W, \vec{b}) = W\vec{x} + \vec{b}, \qquad (4)$$

$$W \in \mathbb{R}^{K \times n}, \vec{b} \in \mathbb{R}^{K}$$
.

- (W, \vec{b}) : parameters of h,
- ▶ W: weights,
- \vec{b} : bias vector.

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in

and Data-Driven
Approach
Image Classification
Challenges in Image
Recognition
Semantic Gap

(5)

Linear Classifier

Approach

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier

Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

Softmax Update
Rule
Prediction and

Parameters Learning



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Image Classification Challenges in Image Recognition Semantic Gap Data-Driven Approach

Training Data Hypothesis Function Hypothesis of Linear Classifier

Understanding of

Hypothesis of Softmax Classifier Loss Function Cross Entropy Loss Regularization

Softmax Update Rule Prediction and

Goal

Parameters Learning



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Goal

▶ Set (W, \vec{b}) to let the score computed match the ground truth,

Image Classification Challenges in Image Recognition Semantic Gap Approach

Training Data Hypothesis of Linear Classifier

Understanding of

Loss Function Cross Entropy Loss Regularization

Softmax Update Rule Prediction and

Parameters Learning



Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Classification Challenges in Image Approach

Hypothesis of Linear Classifier

Understanding of Loss Function Cross Entropy Loss Regularization

Softmax Update Rule Prediction and

Goal

- ▶ Set (W, \vec{b}) to let the score computed match the ground truth.
- ▶ Once learning done, we can throw away the training set, and make prediction based on (W, \vec{b}) .

Understanding of Hypothesis Function



$$h(\vec{x}; W, \vec{b}) = W\vec{x} + \vec{b}$$

$$= \begin{bmatrix} \vec{w}_0^T \\ \vec{w}_1^T \\ \vdots \\ \vec{w}_{K-1}^T \end{bmatrix} \vec{x} + \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{K-1} \end{bmatrix}$$

$$= \begin{bmatrix} \vec{w}_0^T \vec{x} + b_0 \\ \vec{w}_1^T \vec{x} + b_1 \\ \vdots \\ \vec{w}_K^T \cdot \vec{x} + b_{K-1} \end{bmatrix}.$$

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

(6)

(8)

Approach
Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear

Hypothesis of Linear Classifier
Understanding of Hypothesis Function
Hypothesis of Softmax Classifier Loss Function Cross Entropy Loss

Regularization
Optimization
Gradient Descent
Softmax Update
Rule

Prediction and

Understanding of Hypothesis Function



$$h(\vec{x}; W, \vec{b}) = W\vec{x} + \vec{b}$$

$$= \begin{bmatrix} \vec{w}_0^T \\ \vec{w}_1^T \\ \vdots \\ \vec{w}_{K-1}^T \end{bmatrix} \vec{x} + \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{K-1} \end{bmatrix}$$

$$= \begin{bmatrix} \vec{w}_0^T \vec{x} + b_0 \\ \vec{w}_1^T \vec{x} + b_1 \\ \vdots \\ \vec{w}_{K-1}^T \vec{x} + b_{K-1} \end{bmatrix}.$$

Simultaneously has K classifiers,

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

(6)

(8)

Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier

Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization
Ontimization

Softmax Update Rule

Understanding of Hypothesis Function



$$h(\vec{x}; W, \vec{b}) = W\vec{x} + \vec{b}$$

$$= \begin{bmatrix} \vec{w}_0^T \\ \vec{w}_1^T \\ \vdots \\ \vec{w}_{K-1}^T \end{bmatrix} \vec{x} + \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{K-1} \end{bmatrix}$$

$$= \begin{bmatrix} \vec{w}_0^T \vec{x} + b_0 \\ \vec{w}_1^T \vec{x} + b_1 \\ \vdots \end{bmatrix} .$$

$$(8)$$

- ► Simultaneously has *K* classifiers,
- ► For each $0 \le d \le K 1$, $\vec{w}_d^T \vec{x} + b_d$ computes the confidence or score \vec{x} belonging to the *d*-th class.

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven

Linear Classifier

Approach

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier

Classifier
Understanding of Hypothesis Function
Hypothesis of Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

71 Softmax Update Rule

Hypothesis of Softmax Classifier



 $\vec{s} = h(\vec{x}) = W\vec{x} + \vec{b}. \tag{9}$

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

inear Classifier

Training Data Hypothesis Function Hypothesis of Linear Classifier Understanding of

Hypothesis of Softmax Classifier Loss Function

Cross Function
Cross Entropy Loss
Regularization
Optimization
Gradient Descent

Softmax Update Rule

Hypothesis of Softmax Classifier



Deep Learning: Why and What

$$\vec{s} = h(\vec{x}) = W\vec{x} + \vec{b}. \tag{9}$$

Probability interpretation

$$\Pr(y = k | \vec{x}) = \frac{\exp(s_k)}{\sum_{j=0}^{K-1} \exp(s_j)} \in [0, 1], \forall k.$$
 (10)

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Approach
Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of

Hypothesis Function

Hypothesis of
Softmax Classifier
Loss Function

Loss Function
Cross Entropy Loss
Regularization
Optimization
Gradient Descent

Softmax Update
Rule
Prediction and



▶ Given h, for each $\vec{x}^{(i)}$, we can compute scores $\vec{s}^{(i)}$ belonging to each class,

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in

Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of

Hypothesis of Softmax Classif

Loss Function Cross Entropy Loss

Regularization
Optimization
Gradient Descent
Softmax Update
Rule
Prediction and



- ► Given h, for each $\vec{x}^{(i)}$, we can compute scores $\vec{s}^{(i)}$ belonging to each class,
- We need a measure $\bar{s}^{(i)}$ does or does not match ground truth $y^{(i)}$,

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of

Hypothesis of Softmax Classif Loss Function

Loss Function Cross Entropy Loss

Regularization
Optimization
Gradient Descent
Softmax Update
Rule
Prediction and



- ► Given h, for each $\vec{x}^{(i)}$, we can compute scores $\vec{s}^{(i)}$ belonging to each class,
- We need a measure $\bar{s}^{(i)}$ does or does not match ground truth $y^{(i)}$,
- ▶ This measure tells us goodness of parameters (W, \vec{b}) ,

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai Challenges in

Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of

Hypothesis of Softmax Classif Loss Function

Loss Function Cross Entropy Loss Regularization

Optimization
Gradient Descent
Softmax Update
Rule
Prediction and



- ► Given h, for each $\vec{x}^{(i)}$, we can compute scores $\vec{s}^{(i)}$ belonging to each class,
- We need a measure $\bar{s}^{(i)}$ does or does not match ground truth $y^{(i)}$,
- ightharpoonup This measure tells us goodness of parameters (W, \vec{b}) ,
- Loss function measures how much $\bar{s}^{(i)}$ does not match $y^{(i)}$.

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai Challenges in

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data Hypothesis Function Hypothesis of Linear Classifier Understanding of

Hypothesis of Softmax Classi Loss Function

Cross Function
Cross Entropy Loss
Regularization
Optimization
Gradient Descent
Softmax Update
Rule

Cross Entropy Loss



Loss function of softmax classifier can be computed by maximum likelihood estimate

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of

Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

Gradient Descent Softmax Update Rule

Cross Entropy Loss



Loss function of softmax classifier can be computed by maximum likelihood estimate

$$W^{\star}, \vec{b}^{\star} = \arg \max_{W, \vec{b}} \prod_{i=1}^{m} p(y^{(i)} | \vec{x}^{(i)}; W, \vec{b})$$
 (11)

$$= \arg\min_{W,\vec{b}} \frac{1}{m} \sum_{i=1}^{m} \left(-s_{y^{(i)}}^{(i)} + \log(\sum_{j=0}^{K-1} \exp(s_{j}^{(i)})) \right) 12)$$

$$\stackrel{\text{def}}{=} \arg\min_{W,\vec{b}} \frac{1}{m} \sum_{i=1}^{m} \operatorname{err}(W, \vec{b}; \vec{x}^{(i)}, y^{(i)})$$
 (13)

$$\stackrel{\text{def}}{=} \arg \min_{W, \vec{b}} J(W, \vec{b}). \tag{14}$$

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier

Loss Function
Cross Entropy Loss
Regularization

Optimization
Gradient Descent
Softmax Update
Rule

Cross Entropy Loss



Loss function of softmax classifier can be computed by maximum likelihood estimate

$$W^{*}, \vec{b}^{*} = \arg \max_{W, \vec{b}} \prod_{i=1}^{m} p(y^{(i)} | \vec{x}^{(i)}; W, \vec{b})$$
 (11)

$$= \arg\min_{W,\vec{b}} \frac{1}{m} \sum_{i=1}^{m} \left(-s_{y^{(i)}}^{(i)} + \log(\sum_{j=0}^{K-1} \exp(s_{j}^{(i)})) \right) 12)$$

$$\stackrel{\text{def}}{=} \arg\min_{W,\vec{b}} \frac{1}{m} \sum_{i=1}^{m} \operatorname{err}(W, \vec{b}; \vec{x}^{(i)}, y^{(i)})$$
 (13)

$$\stackrel{\text{def}}{=} \arg\min_{W \ \vec{b}} J(W, \vec{b}). \tag{14}$$

For simplicity, write $err(W, \vec{b}; \vec{x}^{(i)}, y^{(i)})$ as $err^{(i)}$.

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier

Loss Function
Cross Entropy Loss
Regularization

Optimization
Gradient Descent
Softmax Update
Rule
Prediction and



Good (W, \vec{b}) min.

$$J(W, \vec{b}) = \frac{1}{m} \sum_{i=1}^{m} -\log \frac{\exp(s_{y^{(i)}}^{(i)})}{\sum_{j=0}^{K-1} \exp(s_{j}^{(i)})}.$$
 (15)

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Classification Challenges in Image Semantic Gap Approach

Training Data Hypothesis of Linear Understanding of Softmax Classifier Loss Function

Cross Entropy Loss Regularization

Softmax Update Rule

Gradient Descent Prediction and



Good (W, \vec{b}) min.

$$J(W, \vec{b}) = \frac{1}{m} \sum_{i=1}^{m} -\log \frac{\exp(s_{y^{(i)}}^{(i)})}{\sum_{j=0}^{K-1} \exp(s_{j}^{(i)})}.$$
 (15)

But that (W, \vec{b}) is not unique

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Approach
Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

_inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss

Regularization
Optimization

Softmax Update Rule



Good (W, \vec{b}) min.

$$J(W, \vec{b}) = \frac{1}{m} \sum_{i=1}^{m} -\log \frac{\exp(s_{y^{(i)}}^{(i)})}{\sum_{j=0}^{K-1} \exp(s_{j}^{(i)})}.$$
 (15)

But that (W, \vec{b}) is not unique

► Suppose (W^*, \vec{b}^*) is the opt. sol.,

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Classification Challenges in Image Semantic Gap

Approach

Training Data Hypothesis of Linear Understanding of Hypothesis of

Loss Function Cross Entropy Loss Regularization

> Softmax Update Rule Prediction and



Good (W, \vec{b}) min.

$$J(W, \vec{b}) = \frac{1}{m} \sum_{i=1}^{m} -\log \frac{\exp(s_{y^{(i)}}^{(i)})}{\sum_{j=0}^{K-1} \exp(s_{j}^{(i)})}.$$
 (15)

But that (W, \vec{b}) is not unique

- ► Suppose (W^*, \vec{b}^*) is the opt. sol.,
- ► Then the result of adding const. c, c > 0 to every element of (W^*, \vec{b}^*) is also opt.,

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai
Challenges in

and Data-Driven
Approach
Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven

Linear Classifier

Approach

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function

Cross Entropy Loss
Regularization
Optimization

Softmax Update Rule



Good (W, \vec{b}) min.

$$J(W, \vec{b}) = \frac{1}{m} \sum_{i=1}^{m} -\log \frac{\exp(s_{y^{(i)}}^{(i)})}{\sum_{j=0}^{K-1} \exp(s_{j}^{(i)})}.$$
 (15)

But that (W, \vec{b}) is not unique

- ► Suppose (W^*, \vec{b}^*) is the opt. sol.,
- ► Then the result of adding const. c, c > 0 to every element of (W^*, \vec{b}^*) is also opt.,
- We handle this problem by make preference of (W, \vec{b}) , this is achieved by adding regularization term $\Omega(W)$,

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai Challenges in

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function

Cross Entropy Loss

9 Regularization Optimization Gradient Descent Softmax Update

Rule Prediction and



▶ Commonly used regularization term is ℓ_2 norm

$$\Omega(W) = \|W\|_F^2 = \sum_{d=0}^{K-1} \sum_{i=0}^{M-1} W_{dj}^2.$$
 (16)

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Classification Challenges in Image Semantic Gap Approach

Training Data Hypothesis of Linear Understanding of Hypothesis of

Loss Function Cross Entropy Loss Regularization

> Softmax Update Rule



▶ Commonly used regularization term is ℓ_2 norm

$$\Omega(W) = \|W\|_F^2 = \sum_{d=0}^{K-1} \sum_{j=0}^{n-1} W_{dj}^2.$$
 (16)

 \triangleright ℓ_2 prefers small and diffused weights, which is good for generalization.

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Classification Challenges in Image Semantic Gap Approach

Training Data Hypothesis of Linear Understanding of Hypothesis of Loss Function

Cross Entropy Loss

Regularization Softmax Update Prediction and

Rule



▶ Commonly used regularization term is ℓ_2 norm

$$\Omega(W) = \|W\|_F^2 = \sum_{d=0}^{K-1} \sum_{j=0}^{n-1} W_{dj}^2.$$
 (16)

 \blacktriangleright ℓ_2 prefers small and diffused weights, which is good for generalization.

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss

Optimization Gradient Descent Softmax Update Rule

Regularization



▶ Commonly used regularization term is ℓ_2 norm

$$\Omega(W) = \|W\|_F^2 = \sum_{d=0}^{K-1} \sum_{i=0}^{n-1} W_{dj}^2.$$
 (16)

 \blacktriangleright ℓ_2 prefers small and diffused weights, which is good for generalization.

Then the total loss function becomes

$$J(W, \vec{b}) = \frac{1}{m} \sum_{i=1}^{m} -\log \frac{\exp(s_{y^{(i)}}^{(i)})}{\sum_{j=0}^{K-1} \exp(s_{j}^{(i)})} + \frac{\lambda}{2} ||W||_{F}^{2} (17)$$
$$= \frac{1}{m} \sum_{i=1}^{m} \exp(s_{j}^{(i)}) + \frac{\lambda}{2} ||W||_{F}^{2} . \tag{18}$$

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Hypothesis Function Hypothesis of Linear Classifier Understanding of Hypothesis Function Hypothesis of Softmax Classifier Loss Function Cross Entropy Loss

Regularization
Optimization
Gradient Descent
Softmax Update

Optimization



Goal:

$$W^{\star}, \vec{b}^{\star} = \arg\min_{W, \vec{b}} J(W, \vec{b})$$
.

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai
Challenges in

Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

(19)

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier

Loss Function
Cross Entropy Loss
Regularization

21) Optimization Gradient De

Softmax Update
Rule
Prediction and

Optimization



Goal:

$$W^{\star}, \vec{b}^{\star} = \arg\min_{W, \vec{b}} J(W, \vec{b}). \tag{19}$$

This is an unconstrained optimization problem.

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognitior and Data-Driven

Approach
Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of

Regularization
21 Optimization
Gradient Descent

Softmax Update
Rule

Loss Function Cross Entropy Loss

Visualing Loss Function



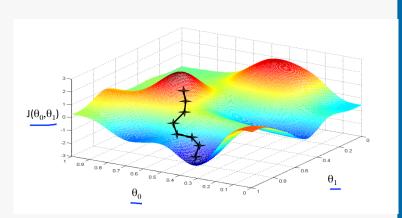


Figure: Visualizing J wrt (W, \vec{b}) is hard, but we can choose 2 directions (θ_0, θ_1) in parameter space, and draw J along that 2 directions. This figure is reproduced from [9].

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss

Regularization
Optimization
Gradient Descent
Softmax Update

Rule
Prediction and



How to find opt. (W, \vec{b})

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Image Classification Challenges in Image Recognition Semantic Gap Data-Driven Approach

Training Data Hypothesis Function Hypothesis of Linear Understanding of

Hypothesis of Softmax Classifier Loss Function Cross Entropy Loss Regularization

Gradient Descent

Softmax Update Rule



How to find opt. (W, \vec{b})

Find them directly is hard,

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Image Classification Challenges in Image Recognition Semantic Gap Data-Driven Approach

Training Data Hypothesis Function Hypothesis of Linear Understanding of

Loss Function Cross Entropy Loss Regularization

Gradient Descent Softmax Update

Rule



How to find opt. (W, b)

- Find them directly is hard,
- Our strategy is start with random (W, \vec{b}) , and refine it iteratively

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Classification Challenges in Image Semantic Gap Approach

Training Data Hypothesis of Linear Understanding of Hypothesis of

Loss Function Cross Entropy Loss Regularization

Gradient Descent Softmax Update



How to find opt. (W, b)

- Find them directly is hard,
- Our strategy is start with random (W, \vec{b}) , and refine it iteratively

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Classification Challenges in Image Semantic Gap Approach

Training Data Hypothesis of Linear Understanding of Hypothesis of

Loss Function Cross Entropy Loss Regularization

Gradient Descent Softmax Update



How to find opt. (W, b)

- Find them directly is hard,
- Our strategy is start with random (W, \vec{b}) , and refine it iteratively

$$W \leftarrow W - \alpha \nabla_W J(W, \vec{b}) = W - \alpha \frac{1}{m} \sum_{i=1}^m \nabla_W \operatorname{err}^{(i)} - \alpha \lambda W,$$
(20)

$$\vec{b} \leftarrow \vec{b} - \alpha \nabla_{\vec{b}} J(W, \vec{b}) = \vec{b} - \alpha \frac{1}{m} \sum_{i=1}^{m} \nabla_{\vec{b}} \operatorname{err}^{(i)}.$$
 (21)

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Image Classification Challenges in Image Semantic Gap

Training Data

Approach

Understanding of Loss Function

Cross Entropy Loss

Regularization Gradient Descent Softmax Update



$$\vec{s}^{(i)} = W\vec{x}^{(i)} + \vec{b}, \forall i, \qquad (22)$$

$$\operatorname{err}^{(i)} = -\log \frac{\exp(s_{y^{(i)}}^{(i)})}{\sum_{j=0}^{K-1} \exp(s_j^{(i)})}.$$
 (23)

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Approach
Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss



Softmax Update Rule



$$\vec{s}^{(i)} = W\vec{x}^{(i)} + \vec{b}, \forall i, \qquad (22)$$

$$\operatorname{err}^{(i)} = -\log \frac{\exp(s_{y^{(i)}}^{(i)})}{\sum_{j=0}^{K-1} \exp(s_j^{(i)})}.$$
 (23)

By caculus,

$$\nabla_{\vec{s}^{(i)}} \operatorname{err}^{(i)} = \frac{\exp(\vec{s}^{(i)})}{\sum_{j=0}^{K-1} \exp(s_j^{(i)})} - \vec{e}_{y^{(i)}}.$$
 (24)

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Approach
Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function

Cross Entropy Loss Regularization

Gradient Descent
Softmax Update
Rule
Prediction and



$$\vec{s}^{(i)} = W\vec{x}^{(i)} + \vec{b}, \forall i, \qquad (22)$$

$$\operatorname{err}^{(i)} = -\log \frac{\exp(s_{y^{(i)}}^{(i)})}{\sum_{j=0}^{K-1} \exp(s_j^{(i)})}.$$
 (23)

By caculus,

$$\nabla_{\vec{s}^{(i)}} \text{err}^{(i)} = \frac{\exp(\vec{s}^{(i)})}{\sum_{j=0}^{K-1} \exp(s_j^{(i)})} - \vec{e}_{y^{(i)}}.$$
 (24)

 $ightharpoonup \vec{e}_{v^{(i)}}$ is a K-dimension vector,

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Image Classification Challenges in Image Semantic Gap Approach

Training Data

Hypothesis of Linear Understanding of Hypothesis of Loss Function Cross Entropy Loss



Rule Prediction and



$$\vec{s}^{(i)} = W\vec{x}^{(i)} + \vec{b}, \forall i, \qquad (22)$$

$$\operatorname{err}^{(i)} = -\log \frac{\exp(s_{y^{(i)}}^{(i)})}{\sum_{i=0}^{K-1} \exp(s_i^{(i)})}.$$
 (23)

By caculus,

$$\nabla_{\vec{s}^{(i)}} \text{err}^{(i)} = \frac{\exp(\vec{s}^{(i)})}{\sum_{i=0}^{K-1} \exp(s_i^{(i)})} - \vec{e}_{y^{(i)}}. \tag{24}$$

- $ightharpoonup \vec{e}_{v^{(i)}}$ is a K-dimension vector,
- it has 1 in $y^{(i)}$'s position, and 0 otherwise.

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Image Classification Challenges in Image Semantic Gap Approach

Training Data

Understanding of Hypothesis of Loss Function Cross Entropy Loss



Prediction and

Softmax Update Rule (cont'd)



By chain rule

$$\nabla_{W} \operatorname{err}^{(i)} = (\nabla_{\vec{s}^{(i)}} \operatorname{err}^{(i)}) \vec{x}^{(i)T}, \qquad (25)$$

$$\nabla_{\vec{b}} \operatorname{err}^{(i)} = \nabla_{\vec{s}^{(i)}} \operatorname{err}^{(i)}. \tag{26}$$

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Image Classification Challenges in Image Semantic Gap

Approach

Training Data Hypothesis of Linear Understanding of Hypothesis of Loss Function Cross Entropy Loss



Softmax Update Rule Prediction and

Prediction and Evaluation



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

_inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of

Understanding of Hypothesis Function Hypothesis of Softmax Classifier Loss Function Cross Entropy Loss Regularization

Softmax Update Rule

Prediction and Evaluation



► Let the classifier model to predict on an unseen test set,

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function

Cross Entropy Loss Regularization Optimization Gradient Descent Softmax Update

Rule

Prediction and Evaluation



- Let the classifier model to predict on an unseen test set,
- ► A good model will have many predictions match the true label.

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai
Challenges in

Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

Softmax Update Rule

Gradient Descent

Prediction



For unseen \vec{x} , the classifier make a prediction by computing

$$\hat{y} = \arg\max_{k} h(\vec{x})_{k} = \arg\max_{k} s_{k}. \tag{27}$$

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

Softmax Update Rule

Prediction



Deep Learning: Why and What

For unseen \vec{x} , the classifier make a prediction by computing

$$\hat{y} = \arg\max_{k} h(\vec{x})_{k} = \arg\max_{k} s_{k}. \tag{27}$$

I.e., assign \vec{x} with label whose score is the highest.

Hao Zhang haomoodzhang@gmai

Image Recognition and Data-Driven Approach

Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Sofomax Classifier
Loss Function
Cross Entropy Loss
Regularization
Optimization

Softmax Update Rule

Evaluation



Given test set

$$D_{test} = \{ (\vec{x}^{(i)}, y^{(i)}) \}_{i=1}^{m_{test}},$$
 (28)

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in

Image Recognitio and Data-Driven Approach Image Classification

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis of
Softmax Classifier
Loss Function

Cross Entropy Loss Regularization

Softmax Update
Rule
Prediction and

Evaluation



Given test set

$$D_{test} = \{ (\vec{x}^{(i)}, y^{(i)}) \}_{i=1}^{m_{test}},$$
 (28)

For each $\vec{x}^{(i)}$, compute prediction $\hat{y}^{(i)}$.

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Approach
Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

Gradient Descent Softmax Update Rule

Evaluation



Given test set

$$D_{test} = \{ (\vec{x}^{(i)}, y^{(i)}) \}_{i=1}^{m_{test}},$$
 (28)

For each $\vec{x}^{(i)}$, compute prediction $\hat{y}^{(i)}$.

Accuracy of the model is

$$Acc = \frac{1}{m_{test}} \sum_{i=1}^{m_{test}} 1\{\hat{y}^{(i)} = y^{(i)}\}.$$
 (29)

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Approach
Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

Table of Contents



Challenges in Image Recognition and Data-Driven Approach

Linear Classifier

Feedforward Neural Networks

Representation Learning
Hypothesis Function
Loss Function
Optimization
Prediction and Evaluation

Convolutional Neural Networks

References

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier

Loss Function Cross Entropy Loss Regularization Optimization Gradient Descent



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognitio and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function

Cross Entropy Loss Regularization Optimization Gradient Descent Softmax Update

Rule Prediction and



Softmax classifier represents a series of linear models,

Why and What

Hao Zhang
haomoodzhang@gmai

Deep Learning:

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

inear Classifier

Training Data Hypothesis Function Hypothesis of Linear Classifier Understanding of Hypothesis Function Hypothesis of Softmax Classifier Loss Function

Cross Entropy Loss Regularization Optimization Gradient Descent Softmax Update

Rule Prediction and



- ► Softmax classifier represents a series of linear models,
- Training is convex optimization,

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in

Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier

Loss Function Cross Entropy Loss Regularization Optimization Gradient Descent



- ► Softmax classifier represents a series of linear models,
- Training is convex optimization,
- Guaranteed to converge to global optimal,

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Ottmax Classifier
Loss Function

Cross Entropy Loss Regularization



- ► Softmax classifier represents a series of linear models,
- Training is convex optimization,
- Guaranteed to converge to global optimal,
- ▶ But it can only use multiple hyperplanes to seperate input space into some simple regions,

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization



- ► Softmax classifier represents a series of linear models,
- Training is convex optimization,
- Guaranteed to converge to global optimal,
- ▶ But it can only use multiple hyperplanes to seperate input space into some simple regions,
- ► For many problems, the raw pixels are not linearly separable.

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Recognition and Data-Driven Approach

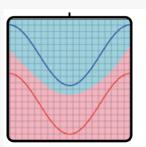
Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization
Ontimization

Feature Engineering





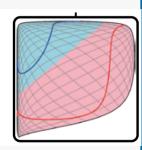


Figure: Find proper representation $\phi(\vec{x})$ to take place \vec{x} , such that $\phi(\vec{x})$'s are linearly separable. This figure is reproduced from [6].

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function

Hypothesis Function Hypothesis of Softmax Classifier Loss Function Cross Entropy Loss

Regularization
Optimization
Gradient Descent
Softmax Update

Rule Prediction and

SIFT Feature







Figure: SIFT can handle occlusion. This figure is reproduced from [7].

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis of Softmax Classifier
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization
Optimization
Gradient Descent
Softmax Undate

Rule Prediction and

HOG feature



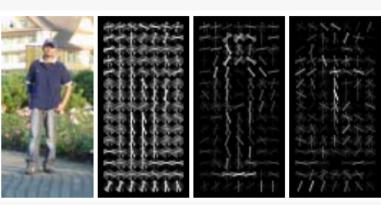


Figure: HOG feature. This figure is reproduced from [1].

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven

Linear Classifie

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization
Optimization



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

_inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss

Regularization Optimization Gradient Descent Softmax Update

Rule Prediction and



• Use fixed ϕ to map \vec{x} to a higher dimension space,

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition

Approach
Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven

Linear Classifier

Approach

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier

Loss Function Cross Entropy Loss Regularization Optimization



- Use fixed ϕ to map \vec{x} to a higher dimension space,
- ► The optimization is still convex,

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier

Loss Function Cross Entropy Loss Regularization Optimization Gradient Descent



- Use fixed ϕ to map \vec{x} to a higher dimension space,
- ► The optimization is still convex,
- ▶ The choice of ϕ is kernel engineering,

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function

Cross Entropy Loss

Regularization
Optimization
Gradient Descent
Softmax Update

Rule Prediction and



Hao Zhang haomoodzhang@gmai

Deep Learning: Why and What

- Image Recognition and Data-Driven Approach
- Image Classification Challenges in Image Recognition Semantic Gap Other Challenges
- Other Challer Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of

Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

Optimization Gradient Des

- Use fixed ϕ to map \vec{x} to a higher dimension space,
- ► The optimization is still convex,
- ▶ The choice of ϕ is kernel engineering,
- ► Commonly used kernel is Gauss kernel: $k(\vec{x}, \vec{x}') = \exp(-\gamma ||\vec{x} \vec{x}'||^2)$.



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

_inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function

Regularization
Optimization
Gradient Descent
Softmax Update
Rule
Prediction and

Cross Entropy Loss



Machine learning can learn the mapping from representation to output.

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Classification Challenges in Image Recognition Semantic Gap Data-Driven Approach

Training Data Hypothesis Function Hypothesis of Linear Understanding of Loss Function

Cross Entropy Loss Regularization Gradient Descent Softmax Update Rule

Prediction and



Machine learning can learn the mapping from representation to output.

Can we also let machine learn appropriate representation?

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition

Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven

Linear Classifier

Approach

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier

Understanding of Hypothesis Function Hypothesis of Softmax Classifier Loss Function Cross Entropy Loss Regularization Optimization Gradient Descent Softmax Update

Rule Prediction and

Machine learning can learn the mapping from representation to output.

Can we also let machine learn appropriate representation? Autoencoder is one example of representation learning.

Deep Learning



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognitio and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

_inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function

Cross Entropy Loss

Regularization
Optimization
Gradient Descent
Softmax Update
Rule
Prediction and

Deep Learning



Learn appropriate mapping directly from raw pixel is still very hard.

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function

Cross Entropy Loss

Regularization
Optimization
Gradient Descent
Softmax Update
Rule
Prediction and

Deep Learning



Learn appropriate mapping directly from raw pixel is still very hard.

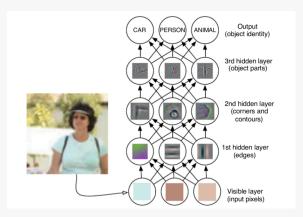


Figure: Deep learning divides it into a series of easy learning problems. $\phi(\vec{x}) = \phi_3(\phi_2(\phi_1(\vec{x})))$. This figure is reproduced from

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven

Linear Classifier

Approach

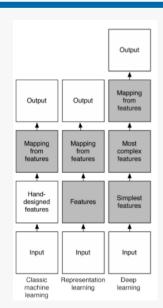
Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis of Softmax Classifier
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization
Optimization
Condicat Descent

Softmax Update Rule

Prediction and

Comparation





Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Challenges in Image Recognitio and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss

Gradient Descent Softmax Update Rule

Regularization

Data Representation



Linear hypothesis function

$$h(\vec{x}) = W\vec{x} + \vec{b}.$$

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in

(30)

Approach
Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven

inear Classifier

Approach

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function

Cross Entropy Loss Regularization

Data Representation



Linear hypothesis function

$$h(\vec{x}) = W\vec{x} + \vec{b}. \tag{30}$$

Changes to

$$h(\vec{x}) = W\phi(\vec{x}) + \vec{b}. \tag{31}$$

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization
Ontimization

Data Representation



Deep Learning: Why and What

Linear hypothesis function

$$h(\vec{x}) = W\vec{x} + \vec{b}. \tag{30}$$

Changes to

$$h(\vec{x}) = W\phi(\vec{x}) + \vec{b}. \tag{31}$$

Nested representation learning

$$\phi(\vec{x}) = \phi_{L-1}(\phi_{L-2}(\dots \phi_2(\phi_1(\vec{x})))). \tag{32}$$

Hao Zhang haomoodzhang@gmai Challenges in

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss

Regularization
Optimization
Gradient Descent
Softmax Update
Rule
Prediction and

Composition of each $\phi_I(\vec{a})$



Linear part

$$\vec{z} = W^{(I)}\vec{a} + \vec{b}^{(I)}, \forall I = 1, 2, \dots, L - 1.$$
 (33)

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognitio

and Data-Driven
Approach
Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges

inear Classifier

Approach

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss

Gradient Descent
Softmax Update
Rule
Prediction and

Regularization

Composition of each $\phi_l(\vec{a})$



Linear part

$$\vec{z} = W^{(I)}\vec{a} + \vec{b}^{(I)}, \forall I = 1, 2, \dots, L - 1.$$
 (33)

Non linear part (activation function)

$$\phi_{I}(\vec{a}) = \max(0, \vec{z}), \forall I = 1, 2, \dots, L - 1.$$

Why and What Hao Zhang haomoodzhang@gmai

Deep Learning:

Image Classification Challenges in Image Semantic Gap Approach

(34)

Training Data Hypothesis of Linear Understanding of

Hypothesis of Loss Function Cross Entropy Loss Regularization

Softmax Update Rule

Prediction and

Activation Function



Rectified linear units, ReLU

$$\max(0,s) = 1\{s > 0\}s. \tag{35}$$

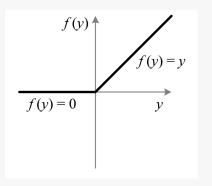


Figure: ReLU. This figure is reproduced from [5].

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

Gradient Descent Softmax Update Rule

Activation Function (cont'd)



Sigmoid

$$\sigma(s) = \frac{1}{1 + \exp(-s)}.\tag{36}$$

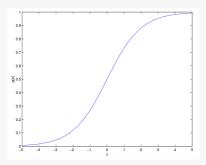


Figure: Sigmoid. This figure is reproduced from [8].

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

In All



If define

$$\vec{a}^{(0)} = \vec{x}.$$

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

(37)

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss

Regularization Optimization

In All



Deep Learning: Why and What

If define

$$\bar{a}^{(0)} = \vec{x}$$
. (37)

$$\vec{z}^{(I)} = W^{(I)} \vec{a}^{(I-1)} + \vec{b}^{(I)}, \forall I = 1, 2, \dots, L-1,$$

$$\vec{a}^{(I)} = \max(0, \vec{z}^{(I)}), \forall I = 1, 2, \dots, L-1,$$
(38)

$$\vec{a}^{(l)} = \max(0, \vec{z}^{(l)}), \forall l = 1, 2, \dots, L - 1,$$
 (39)

$$h(\vec{x}) = W^{(L)} \vec{a}^{(L-1)} + \vec{b}^{(L)}$$
 (40)

Hao Zhang haomoodzhang@gmai

Image Classification Challenges in Image Semantic Gap Approach

Training Data Hypothesis of Linear Understanding of Hypothesis of Loss Function Cross Entropy Loss Regularization

Softmax Update Rule Prediction and

In All



Deep Learning: Why and What Hao Zhang

If define

$$\bar{a}^{(0)} = \vec{x} \,.$$
(37)

$$\bar{z}^{(I)} = W^{(I)} \bar{a}^{(I-1)} + \bar{b}^{(I)}, \forall I = 1, 2, \dots, L-1,$$
(38)

$$\vec{a}^{(I)} = \max(0, \vec{z}^{(I)}), \forall I = 1, 2, \dots, L - 1,$$
 (39)

$$h(\vec{x}) = W^{(L)}\vec{a}^{(L-1)} + \vec{b}^{(L)}.$$
 (40)

 $ightharpoonup \vec{a}^{(L-1)}$ is the representation $\phi(\vec{x})$.

haomoodzhang@gmai Challenges in Image Recognition

Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

Artificial Neural Networks



Deep Learning: Why and What Hao Zhang

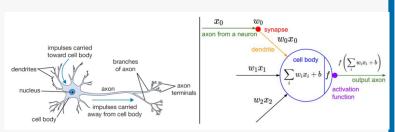


Figure: Biology model and math model. This figure is reproduced from [3].

Challenges in Image Recognition

Approach
Image Classification
Challenges in Image
Recognition
Semantic Gap

Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of

Understanding of Hypothesis Function Hypothesis of Softmax Classifier Loss Function Cross Entropy Loss

Regularization
Optimization
Gradient Descent
Softmax Update

Architecture



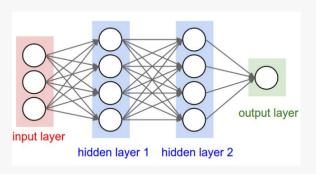


Figure: Multilayer feedforward nueral network. This figure is reproduced from [3].

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai Challenges in

Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier

Cross Entropy Loss Regularization Optimization Gradient Descent Softmax Update

Loss Function

Loss Function



Deep Learning: Why and What

$$J(W, \vec{b}) = \frac{1}{m} \sum_{i=1}^{m} -\log \frac{\exp(s_{y^{(i)}}^{(i)})}{\sum_{j=0}^{K-1} \exp(s_{j}^{(i)})} + \sum_{l=1}^{L} \frac{\lambda}{2} \|W^{(l)}\|_{L^{2}}^{2}$$

$$= \frac{1}{m} \sum_{i=1}^{m} \operatorname{err}^{(i)} + \sum_{l=1}^{L} \frac{\lambda}{2} \|W^{(l)}\|_{F}^{2}.$$
 (42)

$$\bar{s}^{(i)} = h(\bar{x}^{(i)}).$$
(43)

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier Training Data

Hypothesis Function Hypothesis of Linear Classifier Understanding of Hypothesis Function Hypothesis of Softmax Classifier Loss Function

Cross Entropy Loss Regularization Optimization Gradient Descent Softmax Update

Gradient Descent



Update rule

$$W^{(I)} \leftarrow W^{(I)} - \alpha \nabla_{W^{(I)}} J = W^{(I)} - \alpha \frac{1}{m} \sum_{i=1}^{m} \nabla_{W^{(I)}} \operatorname{err}^{(i)} - \alpha \lambda W^{(I)},$$

(44)

$$\vec{b}^{(I)} \leftarrow \vec{b}^{(I)} - \alpha \nabla_{\vec{b}^{(I)}} J = \vec{b}^{(I)} - \alpha \frac{1}{m} \sum_{i=1}^{m} \nabla_{\vec{b}^{(I)}} \operatorname{err}^{(i)}, \forall I. \quad (45)$$

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges

Linear Classifier Training Data

Approach

Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

Gradient Descent (cont'd)



Deep Learning: Why and What Hao Zhang

$$\nabla_{\vec{s}^{(i)}} \operatorname{err}^{(i)} = \frac{\exp(\vec{s}^{(i)})}{\sum_{j=0}^{K-1} \exp(s_j^{(i)})} - \vec{e}_{y^{(i)}},$$
 (46)

$$\bar{s}^{(i)} = h(\vec{x}^{(i)}) = W^{(L)}\bar{a}^{(L-1)} + \vec{b}^{(L)}.$$

haomoodzhang@gmai

Image Classification Challenges in Image Semantic Gap Approach

(47)

Training Data Hypothesis of Linear Understanding of Hypothesis of Loss Function Cross Entropy Loss Regularization

Softmax Update Rule Prediction and

Gradient Descent (cont'd)



$$\nabla_{\vec{s}^{(i)}} \operatorname{err}^{(i)} = \frac{\exp(\vec{s}^{(i)})}{\sum_{j=0}^{K-1} \exp(s_j^{(i)})} - \vec{e}_{y^{(i)}},$$
 (46)

$$\vec{s}^{(i)} = h(\vec{x}^{(i)}) = W^{(L)}\vec{a}^{(L-1)} + \vec{b}^{(L)}$$
.

Easy one

$$\nabla_{W^{(L)}}\operatorname{err}^{(i)} = (\nabla_{\vec{s}^{(i)}}\operatorname{err}^{(i)})\vec{x}^{(i)T}, \qquad (48)$$

$$\nabla_{\vec{c}^{(I)}}\operatorname{err}^{(i)} = \nabla_{\vec{c}^{(I)}}\operatorname{err}^{(i)}. \qquad (49)$$

$$\nabla_{\vec{b}^{(L)}} \operatorname{err}^{(i)} = \nabla_{\vec{s}^{(i)}} \operatorname{err}^{(i)}.$$

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Image Classification Challenges in Image Semantic Gap Approach

(47)

Training Data

Hypothesis of Linear Understanding of Hypothesis of

Loss Function Cross Entropy Loss Regularization Softmax Update

Gradient Descent (cont'd)



Deep Learning: Why and What Hao Zhang

 $\nabla_{\vec{s}^{(i)}} \operatorname{err}^{(i)} = \frac{\exp(\vec{s}^{(i)})}{\sum_{j=0}^{K-1} \exp(s_j^{(i)})} - \vec{e}_{y^{(i)}},$ (46)

$$\vec{s}^{(i)} = h(\vec{x}^{(i)}) = W^{(L)}\vec{a}^{(L-1)} + \vec{b}^{(L)}.$$

Easy one

$$\nabla_{W^{(L)}}\operatorname{err}^{(i)} = (\nabla_{\vec{s}^{(i)}}\operatorname{err}^{(i)})\vec{x}^{(i)T}, \qquad (48)$$

$$\nabla_{\vec{c}^{(L)}}\operatorname{err}^{(i)} = \nabla_{\vec{s}^{(i)}}\operatorname{err}^{(i)}. \qquad (49)$$

Hard One: $\nabla_{W^{(I)}} \operatorname{err}^{(I)}$ and $\nabla_{\vec{b}^{(I)}} \operatorname{err}^{(I)}$, I < L?

haomoodzhang@gmai Challenges in Image Recognition

Challenges in Image Recognitior and Data-Driven Approach Image Classification

Challenges in Image

(47)

Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier Training Data

Hypothesis Function Hypothesis of Linear Classifier Understanding of Hypothesis Function Hypothesis of Softmax Classifier Loss Function

Cross Entropy Loss Regularization

Error Back-Propagation



At layer 1

$$\vec{z}^{(I)} = W^{(I)} \vec{a}^{(I-1)} + \vec{b}^{(I)},$$

$$\vec{a}^{(I)} = \max(0, \vec{z}^{(I)}).$$
(50)

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Approach
Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Data-Driven

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss

Regularization

Error Back-Propagation



At layer 1

$$\vec{z}^{(I)} = W^{(I)} \vec{a}^{(I-1)} + \vec{b}^{(I)},$$
 (50)

$$\bar{a}^{(I)} = \max(0, \bar{z}^{(I)}).$$
 (51)

Assume $\nabla_{\vec{a}^{(l)}} \mathrm{err}^{(i)}$ is known, then

$$\nabla_{\vec{z}^{(l)}} \operatorname{err}^{(i)} = \left(\frac{\partial \vec{a}^{(l)}}{\partial \vec{z}^{(l)}}\right)^{T} \nabla_{\vec{a}^{(l)}} \operatorname{err}^{(i)}$$
(52)

$$=1\{\bar{z}^{(l)}>0\}\odot\nabla_{\bar{a}^{(l)}}\mathrm{err}^{(i)}\,. \tag{53}$$

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

and Data-Driven
Approach
Image Classification
Challenges in Image
Recognition
Semantic Gap

near Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier

Approach

Hypothesis Function
Hypothesis of Linea
Classifier
Understanding of
Hypothesis Functio
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

Error Back-Propagation



At layer 1

$$\bar{z}^{(I)} = W^{(I)} \bar{a}^{(I-1)} + \bar{b}^{(I)},$$

$$\bar{a}^{(I)} = \max(0, \bar{z}^{(I)}).$$
(50)

Assume $\nabla_{\vec{a}^{(i)}} \operatorname{err}^{(i)}$ is known, then

$$abla_{ec{z}^{(l)}} \mathrm{err}^{(i)} = (rac{\partial ec{a}^{(l)}}{\partial ec{z}^{(l)}})^T
abla_{ec{a}^{(l)}} \mathrm{err}^{(i)}$$

$$(i) = (\frac{\partial a^{(i)}}{\partial a^{(i)}})^T$$

$$=1\{\vec{z}^{(l)}>0\}\odot\nabla_{\vec{z}^{(l)}}\mathrm{err}^{(i)}.$$

$$\nabla_{\vec{a}^{(l-1)}} \operatorname{err}^{(i)} = \left(\frac{\partial \vec{z}^{(l)}}{\partial \vec{a}^{(l-1)}}\right)^T \nabla_{\vec{z}^{(l)}} \operatorname{err}^{(i)}$$

$$=W^T\nabla_{\vec{z}^{(l)}}\mathrm{err}^{(i)},$$

$$abla_{W^{(l)}}\operatorname{err}^{(i)} = (
abla_{ec{z}^{(l)}}\operatorname{err}^{(i)})ar{s}^{(l-1)T},
abla_{ec{z}^{(l)}}\operatorname{err}^{(i)} =
abla_{ec{z}^{(l)}}\operatorname{err}^{(i)}.$$

(51)

(52)

(53)

(54)

(55)

(56)

(57

Image Classification

Challenges in Image

Semantic Gap

Approach

Training Data

Cross Entropy Loss

Loss Function

Regularization

Softmax Update Prediction and

Understanding of

Hypothesis of

Prediction and Evaluation



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognitio and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier

Loss Function Cross Entropy Loss Regularization

Softmax Update Rule

Prediction and

Prediction and Evaluation



Why and What
Hao Zhang
haomoodzhang@gmai

Deep Learning:

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier

Loss Function Cross Entropy Loss Regularization

Gradient Descent Softmax Update Rule

Prediction and

The same as linear models,

Prediction and Evaluation



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Image Classification Challenges in Image Recognition Semantic Gap Data-Driven Approach

Training Data Hypothesis Function Hypothesis of Linear Understanding of Hypothesis of Loss Function

Cross Entropy Loss Regularization Gradient Descent

Softmax Update Rule Prediction and

The same as linear models,

▶ But $\vec{s}^{(i)} = h(\vec{x}^{(i)})$ has different form.

Table of Contents



Challenges in Image Recognition and Data-Driven Approach

Linear Classifier

Feedforward Neural Networks

Convolutional Neural Networks
Training Data
Hypothesis Function

References

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier

Cross Entropy Loss Regularization Optimization Gradient Descent

Softmax Update Rule

Loss Function

Training Data



Deep Learning:

Require image input

$$D = \{ (\mathbf{X}^{(i)}, y^{(i)}) \}_{i=1}^{m}, \qquad (58)$$

$$\mathbf{X}^{(i)} \in \mathbb{R}^{H \times W \times D}, \forall i \,, \tag{59}$$

$$y^{(i)} \in \{0, 1, 2, 3, \dots, K - 1\}, \forall i.$$
 (60)

Why and What

Hao Zhang
haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss

Regularization
Optimization
Gradient Descent
Softmax Update
Rule
Prediction and

Hypothesis Function



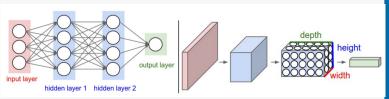


Figure: Normal neural networks and CNN. This figure is reproduced from [3].

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifie

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

Gradient Descent Softmax Update Rule

Expand Fully Connected Affine to Tensor



$$\vec{a} = W\vec{x} + \vec{b}$$

$$= \begin{bmatrix} \vec{w}_0^T \\ \vec{w}_1^T \\ \vdots \\ \vec{w}_{n_l-1}^T \end{bmatrix} \vec{x} + \vec{b}$$

$$= \begin{bmatrix} \vec{w}_0^T \vec{x} + b_0 \\ \vec{w}_1^T \vec{x} + b_1 \end{bmatrix} .$$

$$= (63)$$

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

Expand Fully Connected Affine to Tensor



$$\vec{a} = W\vec{x} + \vec{b}$$

$$= \begin{bmatrix} \vec{w}_0^T \\ \vec{w}_1^T \\ \vdots \\ \vec{w}_{n_l-1}^T \end{bmatrix} \vec{x} + \vec{b}$$

$$= \begin{bmatrix} \vec{w}_0^T \vec{x} + b_0 \\ \vec{w}_0^T \vec{x} + b_0 \end{bmatrix}$$
(62)

$$a_{d_l} = \vec{w}_{d_l}^T \vec{x} + b_{d_l} = \vec{w}_{d_l} \odot \vec{x} + b_{d_l}, \forall d_l.$$
 (64)

Hao Zhang haomoodzhang@gmai Challenges in

Deep Learning: Why and What

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

(63)

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss

Softmax Update Rule

Regularization

Expand Fully Connected Affine to Tensor (cont'd)



Learning And Mining from DatA

Deep Learning:

Why and What Hao Zhang

When $\mathbf{X} \in \mathbb{R}^{H_{l-1} \times W_{l-1} \times D_{l-1}}$.

haomoodzhang@gmai Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier Training Data

Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization
Optimization

Fully Connected Affine to Tensor Expand (cont'd)



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

When $\mathbf{X} \in \mathbb{R}^{H_{l-1} \times W_{l-1} \times D_{l-1}}$. $\mathbf{W}_{d_l} \in \mathbb{R}^{H_{l-1} \times W_{l-1} \times D_{l-1}}$.

$$\mathbf{N}_{d_l} \in \mathbb{R}^{H_{l-1} \times W_{l-1} \times D_{l-1}},$$

Approach

$$a_{d_{l}} = \mathbf{W}_{d_{l}} \odot \mathbf{X} + b_{d_{l}} = \sum_{l=1}^{H_{l-1}-1} \sum_{l=1}^{W_{l-1}-1} \sum_{l=1}^{D_{l-1}-1} \mathbf{X}(i,j,d) \mathbf{W}_{d_{l}}(i,j,d) + b_{d_{l}} \circ \mathbf{X}_{l} \circ$$

(65)

Hypothesis of Linear Understanding of

Loss Function Cross Entropy Loss Regularization

Softmax Update Rule Prediction and

Image Classification Challenges in Image

Two Simplifications of Convolutional Layer



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Image Classification Challenges in Image Recognition Semantic Gap Data-Driven Approach

Training Data Hypothesis Function Hypothesis of Linear Understanding of

Loss Function Cross Entropy Loss Regularization Gradient Descent

Softmax Update Rule Prediction and

Sparse Connection

Parameter Sharing

Sparse Connection



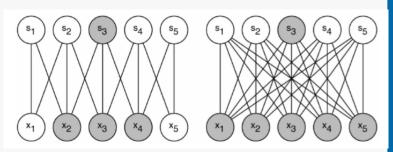


Figure: Sparse connection in 1 dimension. This figure is reproduced from [4].

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai Challenges in Image Recognition

Image Recognition and Data-Driven Approach Image Classification Challenges in Image Recognition

Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifie

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of

Hypothesis Function Hypothesis of Softmax Classifier Loss Function Cross Entropy Loss

Regularization
Optimization
Gradient Descent
Softmax Update

Sparse Connection (cont'd)



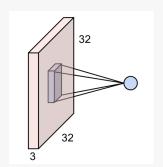


Figure: Sparse connection in 3 dimensions. This figure is reproduced from [3].

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifie

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier

Cross Entropy Loss Regularization Optimization Gradient Descent Softmax Update

Rule Prediction and

Loss Function

Sparse Connection (cont'd)



RF size input image

Figure: Activation map. This figure is reproduced from [2].

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven

Linear Classifier Training Data

Approach

Hypothesis Function Hypothesis of Linear Classifier Understanding of Hypothesis Function Hypothesis of Softmax Classifier Loss Function Cross Entropy Loss

Regularization
Optimization
Gradient Descent
Softmax Update
Rule
Prediction and

Sparse Connection (cont'd)



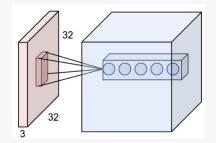


Figure: Input and output tensor. This figure is reproduced from [3].

Deep Learning: Why and What

Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifie

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

Parameter Sharing



Deep Learning: Why and What Hao Zhang

If a motif can appear in one part of the image, it could appear anywhere,

haomoodzhang@gmai

Image Classification Challenges in Image Semantic Gap Data-Driven

Approach

Training Data Hypothesis Function Hypothesis of Linear Understanding of

Loss Function Cross Entropy Loss Regularization Gradient Descent

Softmax Update Rule

Prediction and

Parameter Sharing



If a motif can appear in one part of the image, it could appear anywhere,

Hence units at different locations sharing the same weights

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

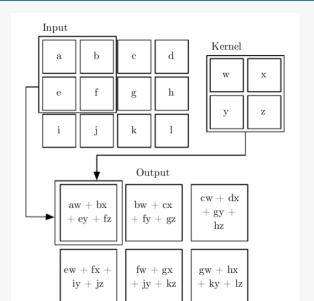
Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function

Cross Entropy Loss Regularization Optimization Gradient Descent Softmax Update

Convolution





Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven

_inear Classifier

Approach

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

Convolution (cont'd)



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

$$\mathbf{A}_{d_{l}} = \mathbf{X} * \mathbf{W}_{d_{l}} + b_{d_{l}}, \forall d_{l}. \tag{66}$$

$$\mathbf{A}(i_l, j_l, d_l) = \sum_{i=0}^{F_1-1} \sum_{j=0}^{F_2-1} \sum_{d=0}^{D_{l-1}-1} \mathbf{X}(i_l+i, j_l+j, d) \mathbf{W}(i, j, d, d_l) + b_{d_l},$$

(67)

- Data-Driven

Image Classification

Challenges in Image

Hypothesis of Linear

Understanding of

Hypothesis of Softmax Classifier

Loss Function Cross Entropy Loss

Regularization

Softmax Update Rule

Prediction and

Pooling Layer



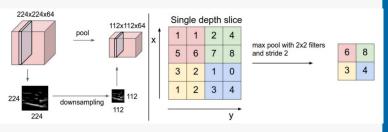


Figure: Max pooling. This figure is reproduced from [3].

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai Challenges in

Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of

Understanding of Hypothesis Function Hypothesis of Softmax Classifier Loss Function Cross Entropy Loss Regularization

Gradient Descent

Pooling Layer (cont'd)



 $\mathbf{A}(i_{l}, j_{l}, d_{l}) = \max_{0 \leq i < F_{1}, 0 \leq j < F_{2}} \mathbf{X}(i_{l}S + i, j_{l}S + j, d_{l}), \forall i_{l}, j_{l}, d_{l}.$ (68)

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven

Approach
Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization

ReLU Layer



 $\mathbf{A}(i_{I},j_{I},d_{I}) = \max(0,\mathbf{X}(i_{I},j_{I},d_{I})), \forall i_{I},j_{I},d_{I}.$ (69)

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai Challenges in

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss

Softmax Update Rule Prediction and

Regularization

Fully Connected Layer



Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in Image Recognitio and Data-Driven

> Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven

inear Classifier

Approach

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier

Loss Function Cross Entropy Loss Regularization

Softmax Update Rule

The same as normal feedforward neural networks.

Table of Contents



Challenges in Image Recognition and Data-Driven Approach

Linear Classifier

Feedforward Neural Networks

Convolutional Neural Networks

References

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai

Challenges in Image Recognition and Data-Driven Approach

Image Classification Challenges in Image Recognition Semantic Gap Other Challenges Data-Driven Approach

inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of

Loss Function Cross Entropy Loss Regularization Optimization Gradient Descent

References I



Navneet Dalal and Bill Triggs.

Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on,

Andrew Ng et al. UFLDL tutorial. Stanford.

Li Fei-Fei, Andrej Karpathy, and Justin Johnson. CS231n: Convolutional neural networks for visual recognition.

Stanford, 2016.

volume 1, pages 886-893. IEEE, 2005.

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in

Image Recognition and Data-Driven Approach

Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization
Optimization

References II



lan Goodfellow, Yoshua Bengio, and Aaron Courville.

Deep learning.

Book in preparation for MIT Press, 2016.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun.

Delving deep into rectifiers: Surpassing human-level performance on imagenet classification.

In Proceedings of the IEEE International Conference on Computer Vision, pages 1026–1034, 2015.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning.

Nature, 521(7553):436–444, 2015.

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai Challenges in

Challenges in Image Recognition and Data-Driven Approach

Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of

Loss Function
Cross Entropy Loss
Regularization
Optimization
Gradient Descent
Softmax Update

References III



David G Lowe.

Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. leee, 1999.

Andrew Ng.
CS229: Machine learning.
Stanford.

Andrew Ng.
Machine learning.
Coursera

Deep Learning: Why and What Hao Zhang haomoodzhang@gmai

Challenges in

Image Recognition and Data-Driven Approach

Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

Linear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier
Loss Function
Cross Entropy Loss
Regularization
Optimization

Thanks



Questions Please!

Hao Zhang haomoodzhang@gmail.com

Deep Learning: Why and What Hao Zhang

haomoodzhang@gmai Challenges in

Image Recognitio and Data-Driven Approach

Image Classification
Challenges in Image
Recognition
Semantic Gap
Other Challenges
Data-Driven
Approach

_inear Classifier

Training Data
Hypothesis Function
Hypothesis of Linear
Classifier
Understanding of
Hypothesis Function
Hypothesis of
Softmax Classifier

Loss Function Cross Entropy Loss Regularization