ECE 884 Deep Learning

Lecture 6: Loss Function

02/04/2021

Logistics

- Google sheet
  - Please form your groups ASAP
- 02/11/2021: Presentations of the Project Ideas
  - Each group has 10mins presentation + 3mins Q&A
  - Please prepare your presentation slides with the following items
    - What topic you aim to work on?
    - What are the existing works?
    - How your idea is different from existing ones?
    - What are the expected results?
    - What is the impact?

Review of last lecture

- Function form#2: parametric linear models
- Function form#3: parametric nonlinear models (DL models)
  - Understand why deep learning models are powerful and importance of massive computation and large data sets.

Today’s lecture

- Loss Function
  - What it is, what is it for?
  - Example: cross-entropy loss function

Recall: Task#1: function form

$$\mathbf{y} = \mathbf{f}(\mathbf{x})$$

output prediction function input

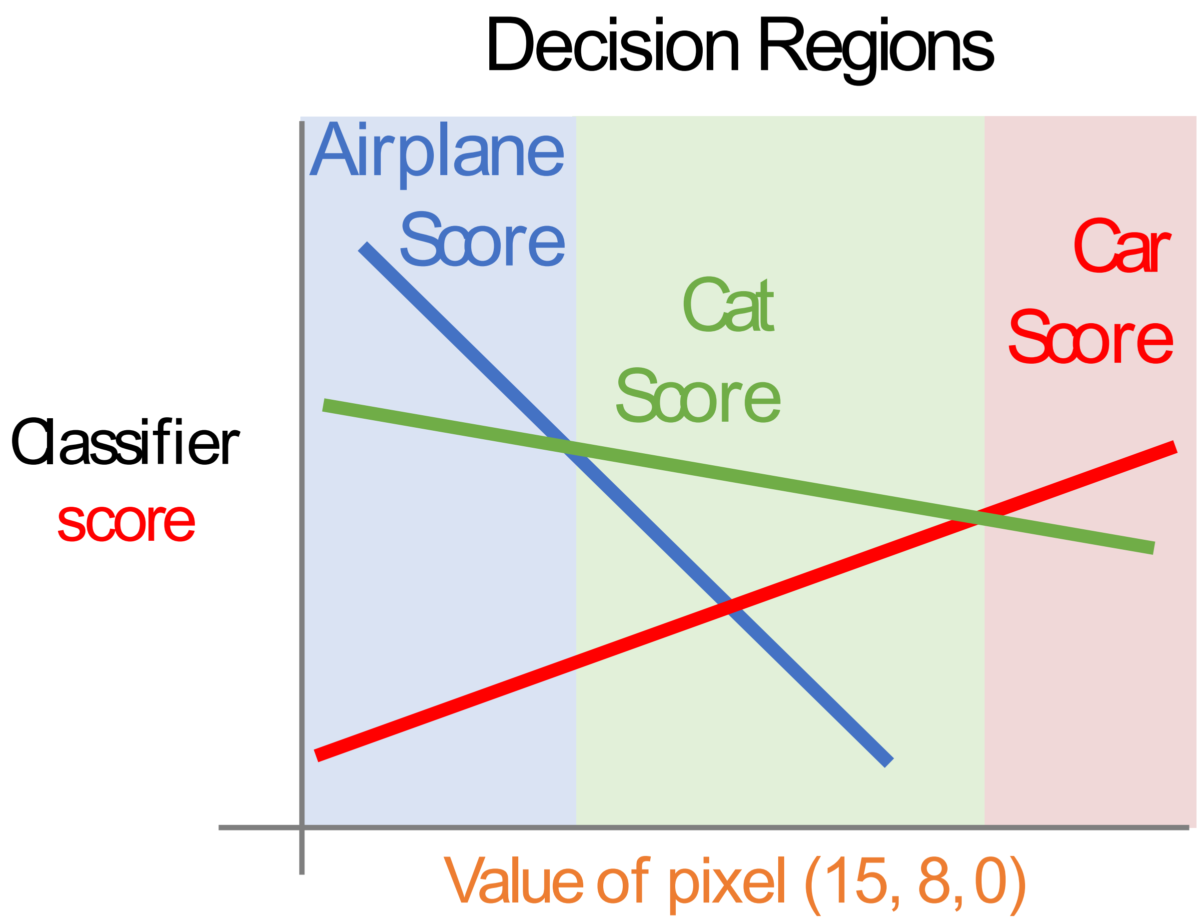
Task#1: function form

Formulation:

- Given training data: $\{(\mathbf{x}_1,y_1), \ldots, (\mathbf{x}_N,y_N)\}$,
- Find $y = f(\mathbf{x}) \in$ " using training data
- such that f is correct on test data

Credit: Lazebnik, Liang

Score Function



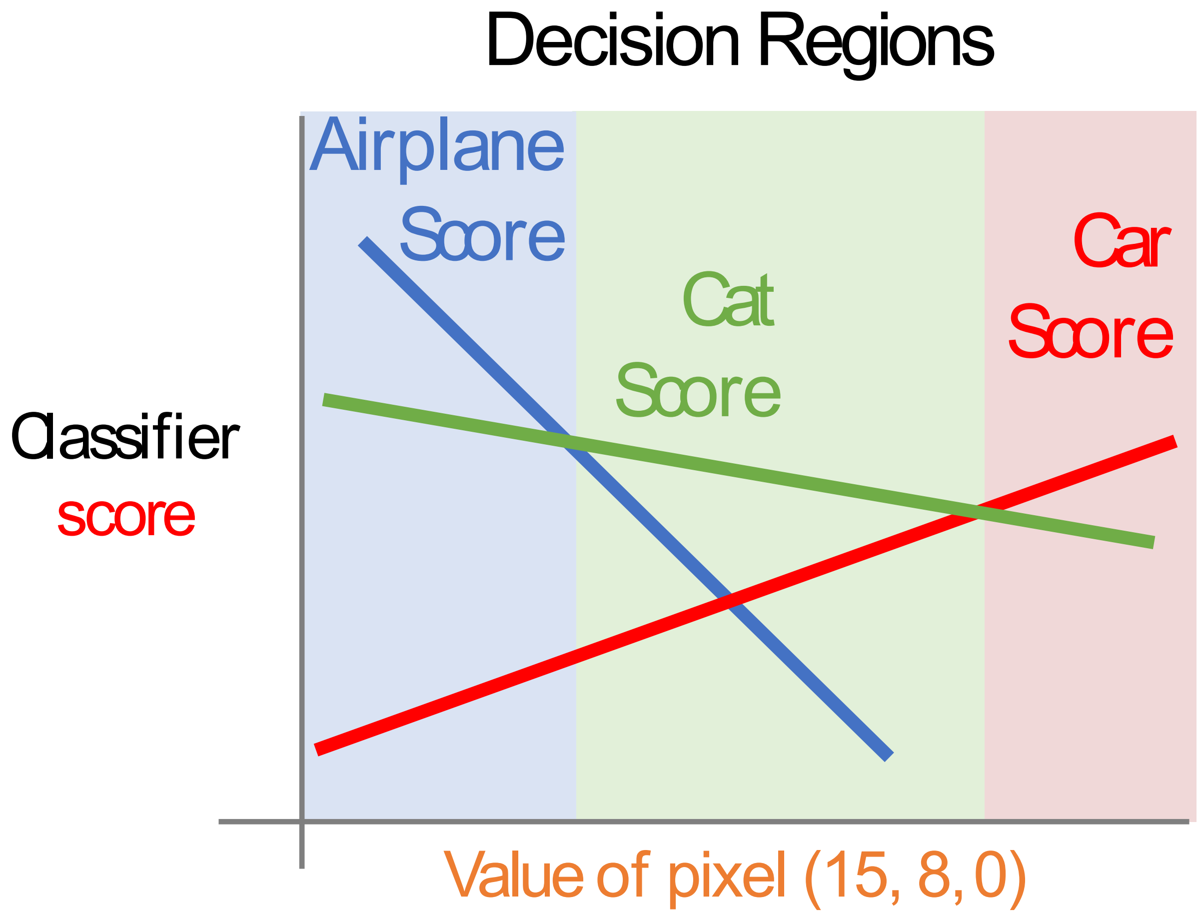
(Linear) Score Function

$$f(x, W) = Wx + b$$

Given a W (and b), we can compute a score for each class given an input x.

Credit: Justin

Score Function



(Linear) Score Function

$$f(x, W) = Wx + b$$

Good classifier: output a **high score** for the **correct** class, and a **low score** for the **wrong** class.

Credit: Justin

Loss Function

What is loss function: a **loss function** is a function that measures how good our classifier is.

Low loss = good classifier
High loss = bad classifier

Credit: Justin

Loss Function

Input

Output

Function form

Loss Function

Input of the loss function

Given a dataset of examples

$$\{(x_i, y_i)\}_{i=1}^{N}$$

Where x_i is image and
y_i is (integer) label

Loss Function

Output: Loss for a single data { xi, yi }

$$L_i(f(x_i, W), y_i)$$

Loss Function

Output: Loss for a single data

$$L_i(f(x_i, W), y_i)$$

Output: Loss for the whole training data set with N data samples

$$L = \frac{1}{N}\sum_i L_i(f(x_i, W), y_i)$$

Loss Function has different forms

Loss Function has different forms

Different problems have different loss functions

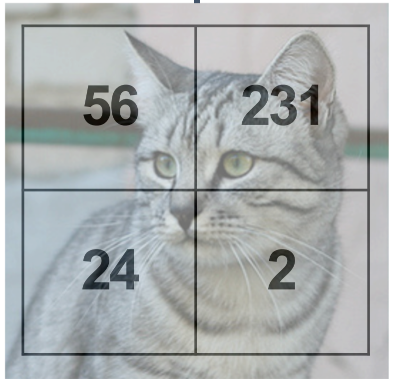
- CV
- NLP
- Speech

Cross-Entropy Loss Function

- One of the most commonly used ones in deep learning
- Nice property: interpret as **probabilities**

Input: Xi

Stretch pixels into column



Input image
Xi

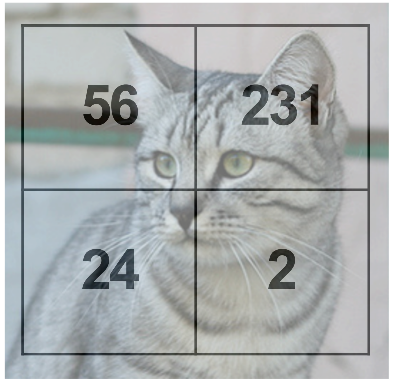
56
231
24
2

Xi

Credit: Justin

Output: Yi

One-hot encoding

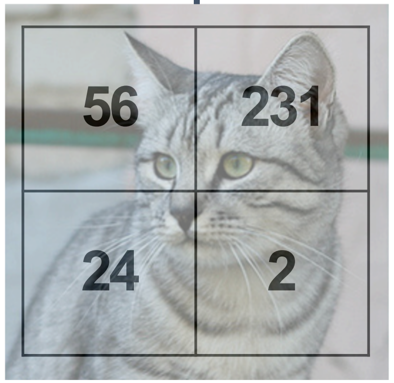


1
0
0

Yi
True Label

Input -> Score Function

Stretch pixels into column



Input image
Xi

$$\underbrace{\begin{bmatrix} 0.2 & -0.5 & 0.1 & 2.0 \\ 1.5 & 1.3 & 2.1 & 0.0 \\ 0 & 0.25 & 0.2 & -0.3 \end{bmatrix}}_{W\ (3,\ 4)} \underbrace{\begin{bmatrix} 56 \\ 231 \\ 24 \\ 2 \end{bmatrix}}_{Xi} + \underbrace{\begin{bmatrix} 1.1 \\ 3.2 \\ -1.2 \end{bmatrix}}_{b\ (3,)} = \underbrace{\begin{bmatrix} 3.2 \\ 5.1 \\ -1.7 \end{bmatrix}}_{(3,)}$$

raw classifier scores (logits)

Credit: Justin

Measure how good the classifier by comparing:

raw classifier scores (logits)

3.2
5.1
-1.7

One-hot encoding

1
0
0

True Label

Measure how good the classifier by comparing:

raw classifier scores (logits)

This is Not probability

3.2
5.1
-1.7

One-hot encoding

1
0
0

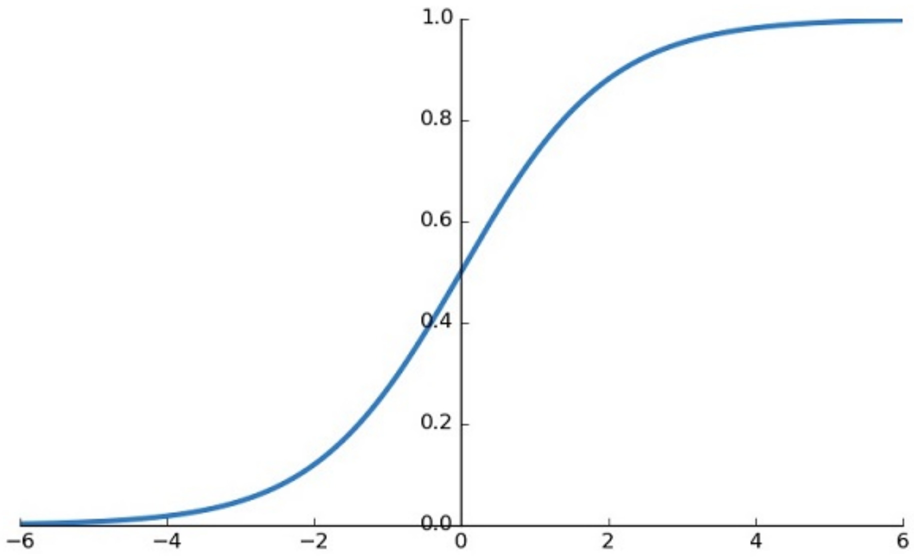
This is probability

True Label

Softmax Transformation

logits

$$P(Y = k \mid X = x_i) = \frac{\exp(s_k)}{\sum_j \exp(s_j)}$$



Softmax Transformation

raw classifier scores (logits)

This is Not probability

$$\begin{bmatrix}3.2\\5.1\\-1.7\end{bmatrix} \xrightarrow{\text{softmax}} \begin{bmatrix}0.12\\0.87\\0.01\end{bmatrix}$$

This is probability

Cross-Entropy:

Measure the distance between 2 probability vectors

Softmax

0.12
0.87
0.01

Cross Entropy

One-hot encoding

1
0
0

True Label

Cross-Entropy:

Measure the distance between 2 probability vectors

Softmax

q

0.12
0.87
0.01

Cross Entropy

$$H(\mathbf{p}, \mathbf{q}) = -\Sigma_i \, p_i \log_2(q_i)$$

One-hot encoding

p

1
0
0

True Label

Cross-Entropy:

Measure the distance between 2 probability vectors

Softmax

q

0.12
0.87
0.01

Cross Entropy

$$H(\mathbf{p}, \mathbf{q}) = -\Sigma_i \, p_i \log_2(q_i)$$

It is not symmetric

One-hot encoding

p

1
0
0

True Label

Cross-Entropy:

Measure the distance between 2 probability vectors

Softmax

q

0.12
0.87
0.01

KL Divergence:

$$D_{KL}(\mathbf{p} \,||\, \mathbf{q}) = H(\mathbf{p}, \mathbf{q}) - H(\mathbf{p})$$

One-hot encoding

p

1
0
0

True Label

Cross-Entropy:

Measure the distance between 2 probability vectors

Softmax

q

0.12
0.87
0.01

Cross Entropy

$$L_i = -\log P(Y = y_i \mid X = x_i)$$

Only with one-hot

One-hot encoding

p

1
0
0

True Label

Cross-Entropy:
Measure the distance between 2 probability vectors

Softmax

q

0.12
0.87
0.01

Cross Entropy

$$L_i = -\log P(Y = y_i \mid X = x_i)$$

Only with one-hot

One-hot encoding

p

1
0
0

True Label

Q: What is the min / max possible loss L_i?

Credit: Justin

Cross-Entropy:

Measure the distance between 2 probability vectors

Softmax

q

0.12
0.87
0.01

Cross Entropy

$$L_i = -\log P(Y = y_i \mid X = x_i)$$

Only with one-hot

One-hot encoding

p

1
0
0

True Label

Q: What is the min / max possible loss L_i?

A: Min 0, max +infinity

Credit: Justin

Cross-Entropy:
Measure the distance between 2 probability vectors

Softmax q
0.12
0.87
0.01

Cross Entropy

$$L_i = -\log P(Y = y_i \mid X = x_i)$$

Only with one-hot

One-hot encoding p
1
0
0

True Label

Q: If all scores are small random values, what is the loss?

Cross-Entropy:

Measure the distance between 2 probability vectors

Softmax

q

0.12
0.87
0.01

Cross Entropy

$$L_i = -\log P(Y = y_i \mid X = x_i)$$

Only with one-hot

One-hot encoding

p

1
0
0

True Label

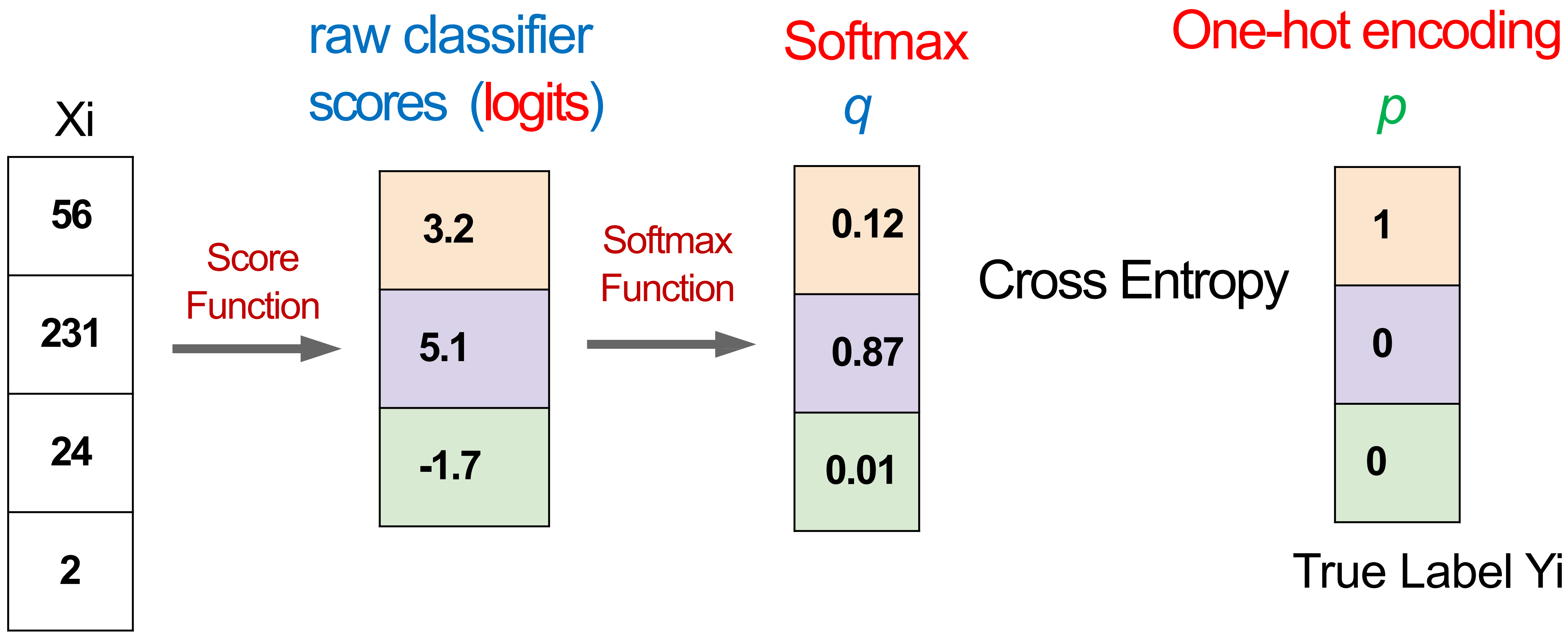
Q: If all scores are small random values, what is the loss?

A: -log(1/C)

log(10) ≈ 2.3

Credit: Justin

Summary


Xi
56
231
24
2
Score
Function
raw classifier
scores (logits)
3.2
5.1
-1.7
Softmax
Function
Softmax
q
0.12
0.87
0.01
Cross Entropy
One-hot encoding
p
1
0
0
True Label Yi

Any Question ?