March 23, 2016

Abstract

1 Overview of Machine Learnnig

1. Logistic Classification
2. Stochastic Optimization
3. Data and Parameter Turning
4. Deep Networks
5. Regularization
6. Convolutional Networks
7. Embeddings
8. Recurrent Models

2 History of Neural Networks

1. Fukushimas Neocognitron - 1980's
2. Lecun's Net -1990
3. Krizhevsky's Alexnet
4. Speech Recognition -2009
5. Computer Vision -2012
6. Machine Translation - 2014

3 Classification

Given set of images and labels in training data. In test data completely new image comes. Classify image. After classification we can do

1. Regression
2. Ranking - In web page. Classify relevant or irrrelevant
3. Reinforcement Learning
4. Detection - Eg : Detect presence or absence of pedestrian

3.1 Logistic Classifier

$$WX + b = Y \tag{1}$$

X - Image Pixels
Y - Labels
W- Weight
b- Bias

3.1.1 Soft Max Function

Softmax function converts scores into probability

$$S(y_i) = \frac{e_i^y}{\sum e_j^y} \tag{2}$$

3.1.2 One Hot Encoding

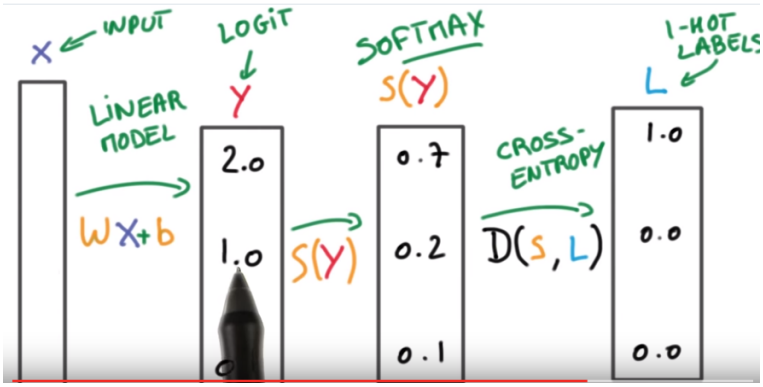
One for correct class and zero to other class labels.

3.1.3 Cross Entrophy

The way to measure distance between two probabilities is called cross entrophy.

$$D(S, L) = -\sum_i L_i log(S_i) \tag{3}$$

Cross Entrophy is not symmetric. $D(S, L) \neq D(L, S)$



4 Training Loss

Loss = Average Cross entrophy. We do to minimize distance between similar labels and maximize distance between dissimilar labels.

$$L = \frac{1}{N}\sum_i D(S(WX_i + b), L_i) \tag{4}$$

5 Gradient Descend

While taking average to calculate training loss we are taking average of distance between probabilities. Gradient descend for two weights is calculated as follows.

$$Gradient Descent for weight w_1 and w_2 = \Delta L(w_1, w_2) \tag{5}$$

But in real-time the weight is computed for all parameters.