Classifiers

Ankur Mishra

1/22/2018, 2/6/2017

Contents

1 Data Driven Approach

- Collect data from internet
- Use ML to train an Model
- Evaluate Model with test data

2 Nearest Neighbor Classifier

- Finds objects that are similar using various functions such as Manhattan/Euclidean Distances and it finds shortest distance from them

- Manahattan (L1) Distance: finds absolute value difference between test and training images
- Euclidean (L2) Distance: finds distance by summing up square of difference between images
- Distances are hyperparameters

2.1 k-Nearest Neighbor

Finds k nearest images, which has a vote for the label, where k is a hyperparametersl

- Non-Parametric - No Defined Number of Parameters
- Never used on images as it has awful test time and distance metrics are unintuitive

3 Linear Classification

3.1 Parametric Approach

Function that outputs W parameters from input x.

$$f(x, W) \tag{1}$$

3.2 Linear Classifier

A linear classifier draws a line and if a point is above the line it is not the class and if it is below the line then it is the class. The value along the line is 0. The line's equation/the score function would be:

$$f(x, W) = Wx + b \tag{2}$$

For an example of an image classifier that has 10 parameters, $f(x, W)$ would be a 10x1 matrix, W would be a 10x3072, and x would be a 3072x1. Bias b and also would be a 10x1 matrix. Matrix multiplication results in the 10x1 matrix.

3.3 Loss Function

Used to find a W which minimizes loss

3.3.1 Muliclass SVM Loss

$$L_i = \sum_{j \neq y_i} max(0, s_j - s_{yi} + 1) \tag{3}$$

j is the object with the highest score. y_i is the object which is trying to be classified. The loss of the function itself is:

$$L = 1/N \sum_{i=1}^{N} L_i \tag{4}$$

3.3.2 Weight Regularization

Weight regulaztion functions are used to favor weights that are more consistent/bring out more features. If I had a vector x which was \$[1,1,1,1], the loss of both scoring vectors \$[1,0,0,0] and $[.25, .25, .25, .25]$ would have the same loss of 1. However, with regularization implemented, the second vector would be favored as it observes more features than the first that only detects the first feature. In general it makes test data better. An equation of a loss function with regularization is:

$$L = 1/N \sum_{i=1}^{N} L_i + \lambda * R(W) \tag{5}$$

λ changes the strength of the regularization

1. L2 Regularization Most common form regularization

$$R(W) = \sum_{k} \sum_{l} W_{k,l}^2 \tag{6}$$

3.3.3 Softmax Classifier

Scores are unnormalized log probabilities of the classes. Probability of class k is:

$$P(Y = k | X = x_i) = (e_k^s)/(\sum_j e_j^s) \tag{7}$$

where $s = f(x_i; W)$

3.3.4 Loss Function

Minimize negative log likelihood of

$$L_i = -log((e_{yi}^s)/(\sum_j e_j^s)) \tag{8}$$

Max is 0, Min is $-\infty$.