CptS 315 - HW 4

Zach Fechko (011711215)

16 November, 2022

Contents

Due: 11:59pm, Sunday, 11/20/2022

Q1

Suppose you are given the following multi-class classification training dta, where each input example has 3 features and output label takes a value from *good*, *bad*, and *ugly*

- $X_1 = (0, 1, 0)$, $y_1 = \text{good}$
- $X_2 = (1, 0, 1)$, $y_2 = \text{bad}$
- $X_3 = (1, 1, 1)$, $y_3 = \text{ugly}$
- $X_4 = (1, 0, 0)$, $y_4 = \text{bad}$
- $X_5 = (0, 0, 1)$, $y_5 = \text{good}$

Suppose we want to learn a linear classifier using multi-class perceptron algorithm and start from the following weights:

- $w_{\text{good}} = (0, 0, 0)$
- $w_{\text{bad}} = (0, 0, 0)$
- $w_{\text{ugly}} = (0, 0, 0)$

Do hand calculations to show how weights change after processing examples in the same order (i.e. one single pass over the 5 training examples)

Q2

Suppose you are given the following binary classification training data, where each input example has three features and output label takes a value *good* or *bad*

- $X_1 = (0,1,0)$, $y_1 = \text{good}$
- $X_2 = (1,0,1)$, $y_2 = \text{bad}$
- $X_3 = (1,1,1)$, $y_3 = \text{good}$
- $X_4 = (1,0,0)$, $y_4 = \text{bad}$
- $X_5 = (0,0,1)$, $y_5 = \text{good}$

Suppose we want to learn a classifier using kernelized perceptron algorithm. Start from the following dual weights:

$$\alpha_1 = 0; \alpha_2 = 0; \alpha_3 = 0; \alpha_4 = 0; \alpha_5 = 0$$

Do hand calculations to show how dual weights change after processing examples in the same order (i.e. one single pass over the 5 training examples). Do this seperately for the following kernels

a. Linear kernel: $K(x, x') = x \cdot x'$
b. Polynomial kernel with degree 3: $K(x, x') = (x \cdot x' + 1)^3$

where $x \cdot x'$ is the dot product of two inputs x and x'. See Algorithm 30. You can ignore the bias term b

Q3

Suppose $x = (x_1, x_2, \ldots, x_d)$ and $z = (z_1, z_2, \ldots, z_d)$ are any two points in a high dimensional space. Suppose you are given the following property, where the right hand side quantity represents the standard Euclidian distance.

$$(\frac{1}{\sqrt{d}}\sum_{i=1}^{d} x_i - \frac{1}{\sqrt{d}}\sum_{i=1}^{d} z_i)^2 \leq \sum_{i=1}^{d}(x_i - z_i)^2$$

We know that the computation of nearest neighbors is very expensive in the high-dimensional space. Discuss how we can make use of the above property to make the nearest neighbors computation efficient?

Q4

We know that we can convert any decision tree into a set of if-then rules, where there is one rule per leaf node. Suppose you are given a set of rules $R = \{r_1, r_2, \ldots, r_k\}$ where r_i corresponds to the i^{th} rule. Is it possible to convert the rule set R into an equivalent decision tree? Explain your construction or give a counter example.

Q5

Read the following two papers and write a brief summary of the main points in at most 3 pages

- Hidden Technical Debt in Machine Learning Systems
- The ML Test score: A Rubric for ML Production Systems