Machine Learning - Homework 3

Mohammad Bahrami - 9724133

December 24, 2021

---

1 Question 1

$$softmax(x_i + c) = \frac{e^{x_i+c}}{\sum_j e^{x_j+c}} = \frac{e^{x_i} \cdot e^c}{\sum_j (e^{x_j} \cdot e^c)} = \frac{e^{x_i} \cdot e^c}{\sum_j (e^{x_j}) \cdot e^c} = \frac{e^{x_i}}{\sum_j (e^{x_j})} = softmax(x_i)$$

---

2 Question 2

2.1 A

If $K = 1$, the first nearest neighbor of the test data is chosen which is from the negative $(-)$ class. Thus, the test data will be assigned to the negative class.

2.2 B

If $K = 3$, two of the nearest neighbors will be from the negative $(-)$ class and one of the neighbors will be from the positive $(+)$ class. As a result, the test data will be assigned to the negative class as the majority of its three nearest neighbors are from the negative class.

2.3 C

If $K > 10$, because there are only five samples from the negative class in the entire dataset, the rest of the neighbors (> 5) will be from the positive class, making the majority of any test data's neighbors from the positive class. Thus, any test data will be assigned to the positive class.

---

3 Question 3

X = (green, 2, tall, no)

$$\hat{y} = \underset{y}{argmax} P(y) \prod_i P(x_i \mid y)$$

$$P(y \mid X) = \frac{P(X \mid y)P(y)}{P(X)} = \frac{\prod_i (P(x_i \mid y)) \cdot P(y)}{P(X)}$$

$$\implies P(M \mid X) = \frac{P(X \mid M)P(M)}{P(X)}$$

$$= \frac{P(color = green \mid M)P(legs = 2 \mid M)P(height = tall \mid M)P(smelly = no \mid M)P(M)}{P(X)}$$

$$= \frac{0.5 \times 0.25 \times 0.25 \times 0.25 \times 0.5}{0.125} = \frac{1}{32}$$

$$\implies P(H \mid X) = \frac{P(X \mid H)P(H)}{P(X)}$$
$$= \frac{P(color = green \mid H)P(legs = 2 \mid H)P(height = tall \mid H)P(smelly = no \mid H)P(H)}{P(X)}$$
$$= \frac{0.25 \times 1 \times 0.5 \times 0.75 \times 0.5}{0.125} = \frac{3}{8}$$

$$\implies P(H \mid X) > P(M \mid X) \implies \hat{y} = H$$

---

4 Question 4

---

5 Question 5

1. Θ_0 is the coefficient of the bias term. which in a linear equation, is the y-intercept. If we put a big regularization coefficient (λ) on it, it will try to reduce the y-intercept of the line. Thus, makes the line be closer to the origin.

2. Θ_1 is the coefficient of the X_1 term. which in a linear equation, represents the line's slope. If we put a big regularization coefficient on it, it will try to reduce the slope of the line and make a horizontal line.

3. Θ_2 is the coefficient of the X_2 term. which in a linear equation, represents the inverse of the line's slope. Thus, a big regularization coefficient on it has the exact opposite effect that it had on Θ_1 and make the slope of the line bigger which produces a vertical line.

---

6 Question 6

As we can see, the error on the train set is decreasing where as it is increasing for the validation set. This is a clear sign of over fitting.

6.1 A

For an over fitted model, adding new parameters will make the condition worse. Adding new parameters to an over fitted model will make the model more prone to over fitting. This makes the bias less but increases the variance.

6.2 B

Adding training data to the model is one of the solutions to fight over fitting. As it shows more examples of data to the model and adds knowledge to it. This makes the bias a little bigger than an over fitted model but doesn't make it too big. Also, this makes the variance smaller as it prevents over fitting.

6.3 C

Increasing the regularization parameter is one of the best ways to prevent a powerful enough but over fitted model from over fitting. If the parameter is well chosen, it will make the bias a little bigger, but decreases the variance dramatically. Although If this parameter is too big, it causes under fitting and makes the bias and variance very big.