ECE/CS/ME 539 – Fall 2024 — Activity 14

1.

(a) Given a logistic regression model with coefficients $\mathbf{w}$ and bias b, how does the model predict the probability of a sample with feature vector $\mathbf{x}$ to belong to the positive class, $P(Y = 1|\mathbf{x})$?

Answer: Given a logistic regression model with coefficients $\mathbf{w}$ and bias b, the probability of a sample with feature vector $\mathbf{x}$ belonging to the positive class $P(Y = 1|\mathbf{x})$ is computed using the logistic function (also known as the sigmoid function).

The logistic regression model first computes a linear combination of the input features and the model's weights:

$$z = \mathbf{w} \cdot \mathbf{x} + b$$

where:

- $\mathbf{w}$ is the weight vector.
- $\mathbf{x}$ is the feature vector.
- b is the bias term.
- $\cdot$ denotes the dot product.

Then, the logistic function is applied to this linear combination to produce a probability value between 0 and 1:

$$P(Y = 1|\mathbf{x}) = \frac{1}{1 + e^{-z}}$$

(b) Normally, a logistic regressor predicts the positive class if $P(Y = 1|\mathbf{x}) > 0.5$ or the negative class otherwise. Suppose that you are building a classifier for a problem, where the cost of a false negative is much larger than the cost of a false positive (e.g., a classifier to diagnose serious medical conditions). Should you increase or decrease the probability threshold p_{thr}?

Answer: We should decrease the probability threshold p_{thr} to a value less than 0.5. By doing so, the classifier becomes more sensitive to potential positive cases, thereby increasing the true positive rate and potentially the false positive rate. This adjustment ensures a reduction in false negatives, albeit at the potential expense of an increase in false positives.

(c) Show that, regardless of the probability threshold p_{thr}, the decision boundary $P(Y = 1|\mathbf{x}) = p_{thr}$ is linear; it defines a hyperplane in the feature space.

Answer: Given a logistic regression model, the probability $P(y = 1|\mathbf{x})$ is defined as:

$$P(y = 1|\mathbf{x}) = \frac{1}{1 + e^{-(\mathbf{w}^T\mathbf{x}+b)}} \tag{1}$$

Where:

- $\mathbf{w}$ is the weight vector.
- $\mathbf{x}$ is the feature vector.
- b is the bias.
- $\mathbf{w}^T\mathbf{x}$ denotes the dot product of $\mathbf{w}$ and $\mathbf{x}$.

For the decision boundary, we set:

$$P(y = 1|\mathbf{x}) = p_{thr} \tag{2}$$

Substituting the expression for $P(y = 1|\mathbf{x})$:

$$\frac{1}{1 + e^{-(\mathbf{w}^T\mathbf{x}+b)}} = p_{thr} \tag{3}$$

Rearranging:

$$1 + e^{-(\mathbf{w}^T\mathbf{x}+b)} = \frac{1}{p_{thr}} \tag{4}$$

$$e^{-(\mathbf{w}^T\mathbf{x}+b)} = \frac{1}{p_{thr}} - 1 \tag{5}$$

Taking the natural logarithm of both sides:

$$-(\mathbf{w}^T\mathbf{x} + b) = \ln\left(\frac{1}{p_{thr}} - 1\right) \tag{6}$$

Rearranging:

$$\mathbf{w}^T\mathbf{x} + b = -\ln\left(\frac{1}{p_{thr}} - 1\right) \tag{7}$$

This equation represents a hyperplane in the feature space. The term $\mathbf{w}^T\mathbf{x}$ is a linear combination of the features, and the right side is a constant determined by p_{thr}. Thus, regardless of the specific value of p_{thr}, the decision boundary remains linear, defining a hyperplane in the feature space.

(d) Suppose each sample is represented by a 2D feature $\mathbf{x} = (x_1, x_2)$. Given a logistic regression model with coefficients $\mathbf{w}$ and bias b, write the equation for the decision boundary in the form $x_2 = mx_1 + c$, where the slope m and intercept c can be computed from $\mathbf{w}, b$, and p_{thr}.

Answer: Given a logistic regression model with a 2D feature vector $\mathbf{x} = [x_1, x_2]$, the linear combination of the features is:

$$z = w_1x_1 + w_2x_2 + b \tag{8}$$

Where:

- w_1 and w_2 are the coefficients corresponding to x_1 and x_2 respectively.
- b is the bias.

For the decision boundary, we have:

$$P(y = 1|\mathbf{x}) = p_{thr} \tag{9}$$

Using the logistic function:

$$p_{thr} = \frac{1}{1 + e^{-z}} \tag{10}$$

Rearranging and taking the natural logarithm:

$$\ln\left(\frac{p_{thr}}{1 - p_{thr}}\right) = w_1x_1 + w_2x_2 + b \tag{11}$$

Isolating x_2 to express the decision boundary in the form $x_2 = mx_1 + c$:

$$x_2 = -\frac{w_1}{w_2}x_1 - \frac{b}{w_2} + \frac{\ln\left(\frac{p_{thr}}{1-p_{thr}}\right)}{w_2} \tag{12}$$

From which we can identify:

$$m = -\frac{w_1}{w_2} \tag{13}$$

$$c = -\frac{b}{w_2} + \frac{\ln\left(\frac{p_{thr}}{1-p_{thr}}\right)}{w_2} \tag{14}$$

Thus, the decision boundary in the desired form is:

$$x_2 = mx_1 + c \tag{15}$$