Week-5, Graded, Solutions

Question-1

Logistic regression is typically used in binary classification settings. The two settings in the options are:

- Predicting whether it will rain on a given day at a given location based on the prevailing weather conditions.
- To predict whether a company's stock price will fall or not on a particular day based on the historical stock market data, and company's annual report.

Question-2

A is a better model than B as the area under the ROC curve is higher for model A than it is for model B. To understand why this makes sense, you can go through the answers in this link.

Question-3

The gradient is given by the following equation:

$$X^T(\sigma(Xw) - y) = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix} \left(\sigma \left(\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix} \right) - \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right)$$

$$= \begin{bmatrix} 0.29 \\ 1.25 \end{bmatrix}$$

Question-4

	True	False
Predicted True	a	b
Predicted False	c	d

- The total number of data-points is:

$$n = a + b + c + d$$

  This is because each point has to fall in one of these four classes: TP, FP, TN, FN.

- The number of data-points that belong to the "True" class is $a + c$.
- The number of data-points that are predicted "False" is equal to $c + d$.

Question-5

If $X = [a, b, c]$, then the polynomial transform of degree 2 on X will result in the following second degree features:

$$a^2, b^2, c^2, ab, bc, ca$$

Adding them to the first degree features and throwing the dummy feature in, we have:

$$[1, a, b, c, a^2, b^2, c^2, ab, bc, ca]$$

Question-6

- Using a polynomial transformation produces a non-linear decision boundary.
- A high-degree polynomial transformation has a lot of capacity, hence it would end up overfitting the training data.

Question-7

The binary cross entropy loss for a set of n examples is given by:

$$\sum_{i=1}^{n} -y^{(i)} \log \hat{y}^{(i)} - (1 - y^{(i)}) \log \left(1 - \hat{y}^{(i)}\right)$$

For the given data-points, this comes out to be 2.398.