Week 5: Revision

Q1. Let D be a random data set generated by the true process

$Y \sim f(X) + \varepsilon$

where; ε is random.

Let $\hat{f}(x_0, D)$ be the prediction for new input, x_0 made by a model trained on D.

Select the true statements.

b. The variance of a KNN model with $K=10$ is usually lower than the variance of a KNN model with $K=1$.

f. The variance of $\hat{f}(X, D)$ at x_0 is $E_D\left(\left(E_D(\hat{f}(x_0, D)) - \hat{f}(x_0, D)\right)^2\right)$

g. $\hat{f}(x_0, D)$ is a random variable.

h. In general, the variance is low for simple models.

Q2. You test a binary-classification algorithm on a given test data set with labels $y \in \{yes, no\}$ and obtain

	True label "yes"	True label "no"
Prediction "yes"	1890	43
Prediction "no"	59	231

Consider a "yes" a positive outcome and compute the specificity of the algorithm.

TP	FP
FN	TN

$$\text{Specificity} = \frac{TN}{(TN+FP)} = \frac{231}{(231+43)} = 0.84$$

Q3. Consider a "yes" as a positive outcome and compute the accuracy of the algorithm.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+FP+FN+TN)} = \frac{1890+231}{1890+43+59+231} = 0.95$$

Q4. Consider the bias and variance of a learning machine, $\hat{f}(X, D)$, trained on the random data set

$$D = \{(X_n, Y_n) \in \mathbb{R}^d \times \mathbb{R}\}, \quad Y_n \sim f(X_n) + \varepsilon_n$$

Assume $\varepsilon_n \sim N(0, \sigma_\varepsilon)$. Then

- The variance of the trained model at $x \in \mathbb{R}^d$ is

$$E_D((\hat{f}(x, D) - E(\hat{f}(x, D)))^2)$$

- The bias of the trained model at $x \in \mathbb{R}^d$ is

$$E_D(\hat{f}(x, D) - f(x))$$

- The expected error of the trained model at $x \in \mathbb{R}^d$ is

$$E_D((\hat{f}(x, D) - f(x))^2)$$

- The irreducible error of the true model at $x \in \mathbb{R}^d$ is

$$E_D(\sigma_\varepsilon)^2 = E_D(Y - f(x))^2$$

- The irreducible error of the true model at $x \in \mathbb{R}^d$ does not depend on $x \in \mathbb{R}^d$ because ε_n does not depened on $X_n \in \mathbb{R}^d$.

- The variance of the trained model at $x \in \mathbb{R}^d$ does depend on $x \in \mathbb{R}^d$ and is a real number.

- The expected value of X_n is $\mu_n = E_D(X_n)$, is a d-dimensional vector.

Q5.

- Suppose you train two prediction algorithms, Algorithm 0 and Algorithm 1, and are looking for evidence that Algorithm 1 produces a better prediction rule than Algorithm 0.
- Your null hypothesis is that the prediction rule produced by Algorithm 1 is not better than the one produced by Algorithm 0.
- You apply the two prediction rules to the same dataset, and the results are summarized in the following contingency table:

	Algorithm 1 correct	Algorithm 1 wrong
Algorithm 0 correct	189	0
Algorithm 0 wrong	5	23

- There is evidence that the prediction rule produced by Algorithm 1 is better. The evidence is statistically significant at level 5%. The evidence is not highly statistically significant (i.e., statistically significant at level 1%).

- Now consider the null hypothesis that the prediction rule produced by the two algorithms have same error probability, the alternative being that their error probabilities are different. In this situation, the same contingency table does not allow you to reject this hypothesis at level 5%. And it does not allow you to reject this hypothesis at level 1%

		A1
	c	w
A0	c 1890	43
	w 59	231

Ques 5

Total observations: 1890 + 43 + 59 + 231 = 2223

0's error rate = (59 + 231) / 2223 = 0.13045

1's error rate = (43 + 231) / 2223 = 0.1232

Q4. $F(x, \lambda) = \lambda_0 + \lambda_1 x$, Mean $= \frac{0+1+2+3}{4} = 1$

$y = mx + c$

$(1,2)$ $(3,4)$

$m = \frac{\Delta y}{\Delta x}$

so,

$F(x_1, \lambda) = \lambda_0 + \lambda_1(0) = \lambda_0$

$F(x_2, \lambda) = \lambda_0 + \lambda_1(1) = \lambda_0 + \lambda_1$

$F(x_3, \lambda) = \lambda_0 + \lambda_1(2) = \lambda_0 + 2\lambda_1$

$F(x_4, \lambda) = \lambda_0 + \lambda_1(3) = \lambda_0 + 3\lambda_1$

$*\ 1 = \lambda_0 + \lambda_1(0)$

Q5. $11 + 5 + 9 + 7 = 32$ $\qquad \frac{21}{32} = 0.65625$

Week 5

Revision

$$* \text{ Accuracy} = \frac{(TP+TN)}{(TP+FP+FN+TN)}$$

Q2. $\cancel{\frac{1890}{2223} = 0.85}$

$$= \frac{(1890 + \cancel{TN}\ 231)}{2223} = 0.95 ✓$$

1890	43
59	231

Q4. Total = 189 + 0 + 5 + 23 = 217

0's = (5+23)/217 = 0.129

1's = 23/217 = 0.105

Q5. $\text{Specificity} = \frac{TN}{(TN+FP)}$

TP	FP
FN	TN

$$= \frac{231}{231+43} = 0.84 ✓$$

189	43
59	231

- Do page 14 more lecture note.
  page 19