Name: _____ NetID: _____

STATISTICS AND DATA SCIENCE 355 / 555

Introductory Machine Learning

Quiz 1 (Practice), Tuesday, October 1, 2019

1. *Logistic regression* (5 points)

   Suppose we have a logistic regression model given by

$$p = \text{logistic}(3 - 0.025 \cdot x_1 - 0.1 \cdot x_2) = \frac{1}{1 + e^{-3 + 0.025 \cdot x_1 + 0.1 \cdot x_2}}$$

   where $p = \mathbb{P}(Y = 1 \,|\, x_1, x_2)$ and $1 - p = \mathbb{P}(Y = 0 \,|\, x_1, x_2)$.

   Consider an input $x = (x_1, x_2) = (100, 10)$. Does the model predict $Y = 1$ or $Y = 0$? Justify your answer.

   Model predicts $Y = 0$, since $\log\left(\frac{p}{1-p}\right) = 3 - .025 \cdot 100 - 0.1 \cdot 10 = 3 - 2.5 - 1 = -.5 < 0$.

2. *Bias and variance* (5 points)

Suppose we have a datapoint $Y \sim N(\theta, 1)$, a single random draw from a Normal distribution with mean $\theta = 5$ and variance 1. Consider an estimator $\widehat{\theta} = bY$ where $0 \leq b \leq 1$.

(a) What is the squared bias of $\widehat{\theta}$?
$(\theta - \mathbb{E}(\widehat{\theta}))^2 = (1-b)^2 5^2$

(b) What is the variance of $\widehat{\theta}$?
$b^2 \operatorname{Var}(Y) = b^2$

Your answers should depend on b.