COSE361: Artificial Intelligence

Homework 3

Spring 2020 - Dr. Hyunwoo J. Kim

Answer the questions in the spaces provided. If you run out of room for an answer, continue on the back of the page.

Name: _____

Student ID: _____

Instructor's name: _____

1. Answer the following questions.
   (a) Explain the Bias-Variance Trade-off.

   (b) Draw a confusion matrix and explain the following concepts.

   (c) Accuracy

   (d) Recall

   (e) Precision

   (f) False Positive

   (g) False Negative

   (h) Explain why both recall and precision need to be considered for evaluation. (hint. trivial predictions to get the best recall or precision)

   (i) Explain 'Occam's Razor' and 'The Curse of Dimensionality'.

   (j) GIve the examples of three random variables that are conditionally independent.

$$X \perp\!\!\!\perp Y | Z$$

(k) (Naïve Classifier) The vocabularies of our spam filter are the following.
V="secret", "offer", "low", "price", "valued", "customer", "today", "dollar", "million", "sports", "is", "for", "play", "healthy", "pizza".

We have the following spam emails for training.
"million dollar offer",
"secret offer today",
"secret is secret"

We have the following normal (non-spam) emails for training.
"low price for valued customer"
"play secret sports today"
"sports is healthy"
"low price pizza"

Using the Naïve Bayes Model, calculate the probabilities.
Write intermediate calculation steps as well.
$P(\text{secret}|\text{spam}) = ?$
$P(\text{spam}) = ?$
$P(\text{sports}|\text{non-spam}) = ?$
$P(\text{dollar}|\text{spam}) = ?$
$P(\text{spam}|\text{“sports is healthy”}) = ?$
$P(\text{non-spam}|\text{“sports is healthy”}) = ?$
$P(\text{“sports is healthy”}|\text{spam}) = ?$
$P(\text{“sports is healthy”}|\text{non-spam}) = ?$