COMS3007 Machine Learning TEST

11 April 2019

This is a closed book test. You may use a calculator.

Time: Two Hours

Question 1. (6 *marks*)

For each of the following scenarios, state whether the problem is a supervised learning problem, unsupervised learning problem, or reinforcement learning problem. Give a one line justification in each case. **1/2 per answer, 1/2 for justification**

(a) Predicting the outcome of an election. **SL - predict a party**

(b) Estimating tomorrow's Bitcoin price. **SL - predict a continuous value**

(c) An autonomous car learning to drive. **RL - learning a behaviour with sparse feedback**

(d) Dividing patients arriving at a hospital into 5 different categories. **UL - no set categories**

(e) Translating documents from Zulu to English. **SL - given mapping between docs**

(f) Segmenting drivers based on their driving styles and behaviours. **UL - no set categories**

Question 2. (4 *marks*)

Explain the difference between overfitting and underfitting. When might each of these occur? **Overfitting - model captures noise in training data (or doesn't generalise well to testing data) (1) when model has too many parameters/is too flexible (1). Underfitting - model isn't able to capture structure in data (1) when model is too inflexible (1).**

Question 3. (3 *marks*)

(a) What do we use training data for? **To learn the model parameters.**

(b) What do we use testing data for? **To report the model quality.**

(c) What do we use validation data for? **To learn the model hyperparameters.**

Question 4. (3 *marks*)

(a) What does it mean for data to be linearly separable? **Classes can be divided by a straight line (1)**

(b) In classification, what is the difference between a generative model and a discriminative model? **Generative model: model the distribution of the data, or class conditional modelling, or model $p(x|y)p(y)$ (1). Discriminative model: model the separation between the classes, or model $p(y|x)$ directly (1)**

Question 5. (3 *marks*)

Bayes' rule is given by $P(y|x) = \frac{P(x|y)P(y)}{P(x)}$.

(a) Which probability in this equation is the *prior*? $P(y)$

(b) Which probability in this equation is the *posterior*? $P(y|x)$

(c) What is the naïve Bayes assumption? **That features are independent given the class, or** $P(x|c) = P(x1|c)P(x2|c)$

Question 6. (11 *marks*)

Consider the training data in Table 1. We now want to classify a new datapoint: (B, C)

TABLE 1. Classification dataset

class	X	Y	Y	X	X	X	Y	Y
feature 1	A	A	B	B	A	A	B	A
feature 2	C	D	D	C	C	D	C	C

(a) Compute $P(X)$ and $P(Y)$. (2) **By counting:** $P(X) = P(Y) = 4/8$. **With any of these results, it doesn't matter if these are simplified.**

(b) Compute $P(B|X)$, $P(B|Y)$, $P(C|X)$, $P(C|Y)$. (4) **By counting:** $P(B|X) = 1/4, P(B|Y) = 2/4, P(C|X) = 3/4, P(C|Y) = 2/4$. **One mark for each.**

(c) Use Naïve Bayes (with Bayes' rule given in the previous question), and the answers in (a) and (b) to compute $P(X|B, C)$ and $P(Y|B, C)$. (4) **First:** $P(B, C|X) = P(B|X)P(C|X) = 0.25 * 0.75 = 0.1875$ **(1)** $P(B, C|Y) = P(B|Y)P(C|Y) = 0.5 * 0.5 = 0.25$ **(1). Then:** $P(X|B, C) = P(B, C|X)P(X)/(0.1875 * 0.5 + 0.25 * 0.5) = 0.1875 * 0.5/0.21875 = 0.4286$ **(1) and** $P(Y|B, C) = P(B, C|Y)P(Y)/0.21875 = 0.25 * 0.5/0.21875 = 0.5714$ **(1)**

(d) What class is (B, C) most likely to be? (1) **By comparing the posteriors, it's Y**

Question 7. (9 *marks*)

We now want to use a decision tree to build a classifier for the same data. Recall that for a feature F and dataset D we define the gain as:

$$Gain(D, F) \quad = \quad H(D) - \frac{1}{|D|} \sum_{f \in values of F} |D_f| H(D_f).$$

Entropy $H(p)$ of a distribution p is $H(p) = -\sum_{i=1}^{n} p_i \log_2(p_i)$, where p_i is the probability of class i.

(a) Compute the entropy of the full dataset $H(D)$. (1) **Entropy** $= -0.5 log 0.5 - 0.5 log 0.5 = 1$

(b) Compute $Gain(D, feature1)$. (3) **Entropy on A:** $H(D_A) = -3/5 log3/5 - 2/5 log2/5 = 0.9710$. **(1) Entropy on B:** $H(D_B) = -1/3 log1/3 - 2/3 log2/3 = 0.9183$. **(1) Gain** $= 1 - 5/8(0.9710) - 3/8(0.9183) = 0.0488$ **(1)**

(c) Compute $Gain(D, feature2)$. (3) **Entropy on C:** $H(D_C) = -3/5 log3/5 - 2/5 log2/5 = 0.9710$. **(1) Entropy on D:** $H(D_D) = -1/3 log1/3 - 2/3 log2/3 = 0.9183$. **(1) Gain** $= 1 - 5/8(0.9710) - 3/8(0.9183) = 0.0488$ **(1)**