COMP5318 Quiz Week 7, Sem 1, 2022

Due No due date **Points** 100 **Questions** 25
Available Apr 4 at 18:00 - Apr 4 at 19:20 about 1 hour **Time Limit** 60 Minutes

Instructions

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	60 minutes	80 out of 100

ⓘ Correct answers are hidden.

Score for this quiz: **80** out of 100
Submitted Apr 4 at 19:00
This attempt took 60 minutes.

Incorrect

Question 1 — 0 / 4 pts

Which of the following statements about Naive Bayes is incorrect?

- Attributes are statistically dependent of one another given the class value.
- Attributes can be nominal or numeric.
- All of the other answers are correct.
- Attributes are equally important.

Question 2

4 / 4 pts

Given the following confusion matrix. What are the precision and recall?

		Predicted Value	
		True	False
Actual Value	True	5	45
	False	5	145

- Precision = 50%; Recall = 50%
- Precision = 10%; Recall = 50%
- Precision = 50%; Recall = 10%
- None of the other options

Question 3

4 / 4 pts

Given the following training data where location, weather, companion and expensive are the features and holiday is the class.

#	Location	Weather	Companion	Expensive	Holiday
1	nice	sunny	annoying	yes	good
2	nice	sunny	annoying	no	bad
3	boring	rainy	great	yes	good
4	boring	sunny	great	yes	bad
5	nice	rainy	great	yes	good
6	boring	rainy	annoying	no	good
7	boring	rainy	great	no	good

What is the value of P(companion=annoying|good)?

- $\frac{2}{5}$
- $\frac{0.6}{P(good)}$
- $\frac{0.4}{P(good)}$
- None of the other options
- $\frac{3}{5}$

Question 4

4 / 4 pts

Given the following training data where location, weather, companion and expensive are the features and holiday is the class.

#	Location	Weather	Companion	Expensive	Holiday
1	nice	sunny	annoying	yes	good
2	nice	sunny	annoying	no	bad
3	boring	rainy	great	yes	good
4	boring	sunny	great	yes	bad
5	nice	rainy	great	yes	good
6	boring	rainy	annoying	no	good
7	boring	rainy	great	no	good

Suppose there's a new unseen example with evidence as follows:

E: location = boring, weather = sunny, companion = annoying, expensive = yes

What is the value of P(bad|E)?

- $\frac{2}{7}$
- $\frac{0.04}{P(E)}$
- $\frac{2}{5}$
- $\frac{0.02}{P(E)}$
- None of the other options

Incorrect

Question 5

0 / 4 pts

Given the following training data where location, weather, companion and expensive are the features and holiday is the class.

#	Location	Weather	Companion	Expensive	Holiday
1	nice	sunny	annoying	yes	good
2	nice	sunny	annoying	no	bad
3	boring	rainy	great	yes	good
4	boring	sunny	great	yes	bad
5	nice	rainy	great	yes	good
6	boring	rainy	annoying	no	good
7	boring	rainy	great	no	good

Suppose there's a new unseen example with evidence as follows:

E: location = boring, weather = sunny, companion = annoying, expensive = yes

What is the value of holiday for E predicted by a Naïve Bayes classifier?

- either of "bad" or "good" with probability of 0.5
- Undetermined due to "zero-frequency" problem or missing value
- holiday = good
- holiday = bad

Question 6

4 / 4 pts

Which of the following statements is true?

- None of the other options
- Leave-one-out cross validation is suitable for large data sets.
- Cross validation can be only used for hyperparameter tuning.
- As k increases, the training time for k-fold cross validation decreases.

Incorrect

Question 7

0 / 4 pts

Given n training points with m features $\boldsymbol{X} \in \mathbb{R}^{n \times m}$ and the corresponding real-value labels $\boldsymbol{y} \in \mathbb{R}^{n}$. What is the closed-form solution of regularized Linear Regression for the given training data?

- $\boldsymbol{w} = (\boldsymbol{X}^T \boldsymbol{X} + \lambda \boldsymbol{I})^{-1} \boldsymbol{X}^T \boldsymbol{y}$, with $\boldsymbol{w} \in \mathbb{R}^{n}$
- $w = (X^T X)^{-1} X^T y$, with $w \in \mathbb{R}^{n}$
- $w = (X^T X + \lambda I)^{-1} X^T y$, with $w \in \mathbb{R}^{m}$
- $w = (X^T X)^{-1} X^T y$, with $w \in \mathbb{R}^{m}$

Question 8

4 / 4 pts

Given a dataset with only 3 data points, each described by 2 features: level, points

A = [2, 8000]

B = [4, 9000]

C = [3, 5000]

After normalisation, what is the value of A, B and C?

- None of other answers
- A = [0.0, 0.75], B = [1, 1], C = [0.5, 0]

A = [0,0.67], B = [1, 1], C = [0.5, 0]

A = [0, 0.25], B = [1, 1], C = [0.5, 0]

Question 9

4 / 4 pts

What statement about the k-Nearest Neighbor algorithm is **incorrect**?

In a two-class domain, k should be an odd number so as to prevent ties.

kNN classifier is a non-linear model.

Irrelevant attributes impair the k-NN classifier's performance

In noisy domains, the 1-NN classifier will outperform k-NN classifier (with k> 1).

Question 10

4 / 4 pts

Given the following training set of 4 examples described by 2 numeric attributes, determine the class of unseen data x = [2,4] using the 3-Nearest Neighbor algorithm.

Note: distance measure is Euclidean distance, k=3

id	Attributes	Label

Id	Attributes	Label
ex1	[1, 3]	pos
ex2	[3, 5]	pos
ex3	[3, 2]	neg
ex4	[5, 2]	neg

- neg
- pos
- pos or neg

Incorrect

Question 11

0 / 4 pts

In a weighted 4-Nearest Neighbour problem to determine the class of data point x, let the distances between x and the 4 nearest neighbours be $d_1 = 1,\ d_2 = 3,\ d_3 = 4,\ d_4 = 5$.

The individual weights are calculated as follows:

$$w_i = \frac{1}{d_i}$$

If the 1st nearest neighbour is positive and the remaining three are negative, then x is classified as

- positive or negative
- negative
- positive

Question 12

4 / 4 pts

Given the closed-form solution of regularized Linear Regression as $w = (X^T X + \lambda I)^{-1} X^T y$, with $X \in R^{n \times m}$. Then its computation complexity is:

- $O(mn^2 + n^3)$
- $O(nk^3 + k^2)$
- **$O(n\,m^2 + m^3)$**
- $O(n\,m^2 + k^3)$

Question 13

4 / 4 pts

Bob wants to use Logistic Regression for binary classification [1,-1]. He obtains a weight vector $w = [0.5, 0.7, -0.2, -0.4]^T$. Now he receives a new sample $x = [1, -1, -2, -1]^T$. If Bob uses the sigmoid function $\sigma\left(w^T x\right)$, then (a) what is the probability of this sample to be classified into class 1; (b) which class the sample will be classified?

Sigmoid function: $\sigma(x) = \frac{1}{1+e^{-x}}$

- (a) 0.19 and (b) -1
- (a) 0.52 and (b) 1
- (a) 0.64 and (b) 1
- (a) 0.35 and (b) -1
- None of other answers

Question 14

4 / 4 pts

Which statement about Gradient Descent (GD) and Stochastic Gradient Descent (SGD) is incorrect?

- SGD requires fewer iterations, which always leads to slower overall convergence.
- SGD reduce computation at each iteration compared to GD.
- Gradient Descent can be used for both convex and non-convex problems.
- If the learning rate is too small, then GD may take a very long time to converge.

Question 15

4 / 4 pts

Alice wants to use the Stochastic Gradient Descent (SGD) update for Ridge Regression with Euclidean norm regularization on the feature matrix $X \in R^{n \times m}$ and target vector $y \in R^n$. Which one below is the correct form?

- $w_{i+1} = w_i - \alpha_i \left[\lambda w_i + (w_i^T x^{(k)} - y^{(k)}) x^{(k)} \right]$

- $w_{i+1} = w_i - \alpha_i \left[\lambda \|w_i\|^2 + \sum_{k=1}^{n} (w_i^T x^{(k)} - y^{(k)}) x^{(k)}\right]$
- $w_{i+1} = w_i - \alpha_i \sum_{k=1}^{n} (w_i^T x^{(k)} - y^{(k)}) x^{(k)}$
- $w_{i+1} = w_i - \alpha_i \left[\lambda w_i + \sum_{k=1}^{n} (w_i^T x^{(k)} - y^{(k)}) x^{(k)}\right]$

Question 16

4 / 4 pts

Which of the following statements about Ridge Regression with regularization term $\lambda||\mathbf{w}||_2^2$ is correct?

- Ridge regression is a non-convex problem.
- The test data can be used to tune λ.
- The Empirical Risk Minimization (ERM) of the above Ridge regression does not have a closed-form solution, but we can use Gradient Descent to obtain the optimal solution.
- Given a large λ, the regression models emphasize minimizing the model complexity than minimizing the training error.

Question 17

4 / 4 pts

What are the advantages of the decision tree?

- A. Decision trees are easy to visualize
- B. Non-linear patterns in the data can be captured easily
- (selected) Both A and B
- None of the other answers

Incorrect

Question 18

0 / 4 pts

A decision tree classifier selects the attribute which has

- (selected) Largest Entropy and largest Information gain
- Smallest Entropy and smallest Information gain
- Smallest Entropy and largest Information gain
- Largest Entropy and smallest Information gain

Question 19

4 / 4 pts

Which method should be used when decision trees suffering from high variance?

- All of the other answers
- Increase the depth of decision trees

- [x] Random forest
- [ ] Reduce training data size

Question 20

4 / 4 pts

Which of the following is an example of a sequential ensemble model?

- [ ] Bagging
- [ ] Random forest
- [ ] Bootstrapping
- [x] AdaBoost

Question 21

4 / 4 pts

Given **n** samples in the dataset **X**. What is the reasonable method to select the number of principal components **k** for preserving **99%** information of original dataset **X**?

- [x] Choose k to be the smallest value so that at least 99% of the variance is retained
- [ ] Choose k to be 99% of n (i.e., k = 0.99 * n, rounded to the nearest integer)

- Choose k to be larger than the number of eigenvalues
- Choose k to be the smallest value so that at least 1% of the variance is retained

Question 22

4 / 4 pts

p1, p2 and p3 are the first, the second and the third principal components, respectively. Which statement is correct?

- Variance along p2 is bigger than variance along p1
- (selected) p1 is orthogonal to p2, p3 is orthogonal to p1 and p2
- p2 is parallel to p3 and p1
- p1 is parallel to p2, p1 is orthogonal to p3

Question 23

4 / 4 pts

In a support vector machine classifier with soft margin which can optionally use kernel trick, when the C parameter is set to infinite, which of the following holds true?

- The soft-margin classifier will not separate the data

SVM models emphasis on maximizing the margin than minimizing the training error

Misclassification would happen

The optimal hyperplane if exists, will be the one that completely separates the data

Question 24

4 / 4 pts

What is the purpose of the Kernel Trick?

To transform the problem from nonlinear to linear

To transform the problem from regression to classification

To transform the data from nonlinearly separable to linearly separable

To transform the problem from supervised to unsupervised learning

Question 25

4 / 4 pts

An image (**m** $\times$ **n** pixels) is compressed by first **k** components with Singular Vector Decomposition (SVD). Which is the correct statement?

We need to store (k$*$m + k + n) matrix elements

- (●) We need to store (k + k∗n + k∗m) matrix elements
- ( ) We need to store (k∗n + k + m) matrix elements
- ( ) We need to store (k∗m + n + 1) matrix elements

Quiz Score: **80** out of 100