Quiz 5

- There are 5 problems total worth 12 points as shown in each question.
- You must not communicate with other students during this test.
- No books, notes allowed.
- No other electronic device except calculators are allowed. You cannot use your mobile as calculators.
- This is a 30 minute exam.
- Do not turn this page until instructed to.
- There are several different versions of this exam.

1. Fill in your information:

Full Name: _____

NetID: _____

Zone 1

1/1. (2 points)

Suppose that we want to select between two prediction models, M1 and M2. We have performed 10 rounds of 10-fold cross-validation on each model, where the same data partitioning in round i is used for both M1 and M2. The error rates obtained for M1 are 30.5, 32.2, 20.7, 20.6, 31.0, 41.0, 27.7, 26.0, 21.5, 26.0. The error rates for M2 are 22.4, 14.5, 22.4, 19.6, 20.7, 20.4, 22.1, 19.4, 16.2, 35.0.

$$Hint: t = \frac{\overline{err}(M_1) - \overline{err}(M_2)}{\sqrt{var(M_1 - M_2)/k}}$$

A. What is the t-statistic between the given two models?

Solution.

$$\overline{err(M_1)} = \frac{30.5 + 32.2 + 20.7 + 20.6 + 31.0 + 41.0 + 27.7 + 26.0 + 21.5 + 26.0}{10} = 27.72$$

$$\overline{err(M_2)} = \frac{22.4 + 14.5 + 22.4 + 19.6 + 20.7 + 20.4 + 22.1 + 19.4 + 16.2 + 35.0}{10} = 21.27$$

$$\overline{var(M1 - M2)} = 1/10 * (30.5 - 22.4 - (27.72 - 21.27))^2 + \cdots) = 68.1$$

$$t = \frac{27.72 - 21.27}{\sqrt{(6.812)}} = 2.47$$

1/2. (2 points)

Suppose that we want to select between two prediction models, M1 and M2. We have performed 10 rounds of 10-fold cross-validation on each model, where the same data partitioning in round i is used for both M1 and M2. The error rates obtained for M1 are 26.0, 21.5, 26.0, 30.5, 32.2, 20.7, 20.6, 31.0, 41.0, 27.7. The error rates for M2 are 19.4, 16.2, 35.0, 22.4, 14.5, 22.4, 19.6, 20.7, 20.4, 22.1.

$$Hint: t = \frac{\overline{err}(M_1) - \overline{err}(M_2)}{\sqrt{var(M_1 - M_2)/k}}$$

A. What is the t-statistic between the given two models?

Solution. same

2/1. (2 points) Imagine a classifier for cancer detection is built by balancing the classes. The classifier creates a new dataset by picking all cancer patients and an equal number of non-cancer patients from the training dataset. The classifier then trains a classifier on the new dataset. Identify one positive and one negative aspect that is associated with the above balancing approach with brief explanation.

Hint: Identify the trade-off.

Solution. +ive:improves sensitivity. new classifier will be better at identifying cancer patients.

-ive:
-prior probabilities are changed therefore poor generalization error.
- classification of negative examples may get very far off because of skipping or poor choice of negative examples.

2/2. (2 points) Imagine a classifier for cancer detection is built by balancing the classes. The classifier creates a new dataset by picking all cancer patients and an equal number of non-cancer patients from the training dataset. The classifier then trains a classifier on the new dataset. Identify any two drawbacks that are associated with the above balancing approach with brief explanation.

Hint: Identify the trade-off.

Solution.

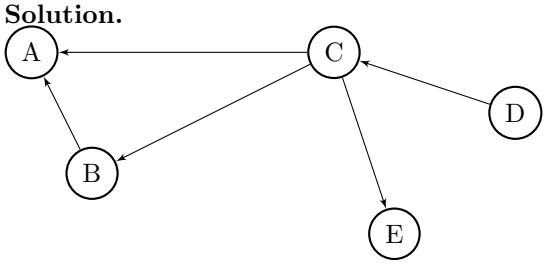
-ive:
-prior probabilities are changed therefore poor generalization error.
- classification of negative examples may get very far off because of skipping or poor choice of negative examples.

3/1. (2 points) Answer the following questions relating to Bayesian belief network.

A. Construct the Bayesian Belief network for the following joint probability distribution.
$P(A, B, C, D, E) = P(A|B, C) \cdot P(B|C) \cdot P(E|C) \cdot P(C|D)$

B. Suppose there are two random variables X and Y, and X causes Y. Briefly explain why a Bayesian belief network cannot be used to learn this causal relationship between X and Y.

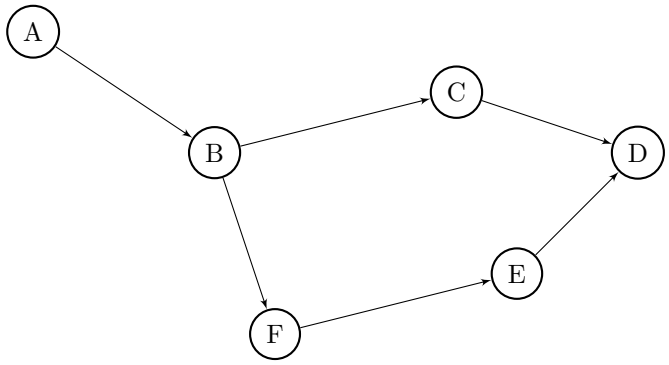
Solution.



- Bayesian belief network captures conditional probabilities or correlation. It is not possible to express causation using these tools.

3/2. (2 points) Answer the following questions relating to Bayesian belief network.

A. Construct the following joint probability distribution for the following Bayesian Belief network.



B. Suppose there are two random variables X and Y, and X causes Y. Briefly explain why a Bayesian belief network cannot be used to learn this causal relationship between X and Y.

Solution.
$P(A, B, C, D, E, F) = P(A) \cdot P(B|A) \cdot P(C|B) \cdot P(F|B) \cdot P(E|F) \cdot P(D|C, E)$
- Bayesian belief network captures conditional probabilities or correlation. It is not possible to express causation using these tools.

4/1. (3 points)
Answer the following questions relating to SVM (Support Vector Machines).

A. Why does a SVM (support vector machines) algorithm have high classification accuracy in high-dimensional space?

B. Why is a SVM algorithm slow in large data sets?

Solution.

A. It is easier to find separating hyperplanes in the same data points are projected into higher dimensions. Therefore higher dimensional space leads to higher classification accuracy.

Alternative, higher dimensions usually encode higher order relationships that allows for better pattern learnt.

B. Training is very slow. The computational complexity of SVM is $O(n^3)$ since it require dot product of the data points. Also projection of data points makes SVM slower leading it to be very slow in larger datasets.

5/1. (3 points)

Assuming that depth of a neural network is defined as maximum path length from input layer to output layer, answer the depth of the following classifiers when treated as neural network.

A. A decision tree.

B. A linear discriminant classifier.

C. An ensemble of RBF (radial basis function) SVMs.

Solution.

A. depth of the decision tree.

B. 1. (just a linear combination of inputs)

C. 3. (An ensemble (can be non-linear) takes one layer, and projection and linear combination of SVM take two layers)

5/2. (3 points) Assuming that depth of a neural network is defined as maximum path length from input layer to output layer, answer the depth of the following classifiers when treated as neural network.

A. A decision tree.

B. An ensemble of linear discriminant classifiers.

C. A RBF (radial basis function) SVM.

Solution.

A. depth of the decision tree.

B. 2. (An ensemble (can be non-linear) takes one layer, a linear combination of inputs takes the other)

C. 2. (projection and linear combination of SVM take two layers)
