

**Quiz 3, Fall 23 (10 points)**

**Section A**

**NAME and EID:** \_\_\_\_\_

**For Multiple Choice Questions, circle the best answer.**

Q1. (2 points) Which of the following is always True for PCA :

- a. **The principal components are the eigenvectors of the data covariance matrix.**
- b. The principal components are the eigenvectors of the data matrix.
- c. Projecting the data into a lower dimensional space using PCA helps improve performance of a predictive model.
- d. The amount of data variance captured in principal components is inversely proportional to the number of components.

Q2. (2 points) Suppose we have a dataset of 500 data points, each having 7 features. The eigenvalues for the first 5 components are 30, 20, 15, 10, and 5. What is the minimum number of principal components needed to explain a variance of at least 50%?

- a. 1
- b. 4
- c. 5
- d. **2**

Q3. (2 points) Which of the following is always true about the data pre-processing methods discussed in the class?

- a. PCA is a non-linear and unsupervised method.
- b. Using an auto-encoder to obtain latent representations of the data is a linear method.
- c. Fisher's Linear Discriminant is an unsupervised method.
- d. **t-SNE is an unsupervised and a non-linear method.**

Q4. (2 points) A fully connected MLP has one input, one hidden layer and one output layer. The number of neurons in the 3 layers are 3, 4 and 2 respectively. How many learnable parameters does this network have in total (including all bias weights)?

- a. **26**
- b. 24
- c. 20
- d. 9

Q5. (2 points) Briefly explain how Huber loss is helpful in dealing with outliers.

In huber loss, large errors are penalized linearly which reduces their impact on the model parameters.

**Quiz 3, Fall 23 (10 points)**

**Section B**

**NAME and EID:** \_\_\_\_\_

**For Multiple Choice Questions, circle the best answer.**

Q1. (2 points) Which of the following statements is always true about PCA?

- a. PCA is susceptible to local optima; trying multiple random initializations may help.
- b. **The directions of projections determined by PCA are orthogonal to one another.**
- c. PCA can be used only to reduce the dimensionality of data by 1 (such as 3D to 2D, or 2D to 1D).
- d. Feature scaling is not useful for PCA, since the eigenvector calculation takes care of this automatically.

Q2. (2 points) State whether the following is an example of feature selection or feature engineering:

Using an autoencoder to obtain a latent space representation of dimension 20 when the original dimension of the data was 75.

- a. **Feature Engineering**
- b. Feature Selection
- c. Both
- d. None of these

Q3. (2 points) Which of the following is FALSE about the data pre-processing methods discussed in the class?

- a. PCA is a linear feature extraction method.
- b. In PCA and t-SNE the number of output features can be specified.
- c. **LASSO produces features that are linear combinations of the original features.**
- d. The PCA objective minimizes the re-construction error along with maximizing the variance of the projected data.

Q4 (2 points) Which of the following is NOT a method for limiting model complexity in an MLP?

- a. **Limiting the number of output units.**
- b. Early stopping the training process.
- c. Limiting the number of hidden units.
- d. Penalizing the L1 norm of weights.

Q5. (2 points) Briefly explain why reducing the number of features via feature selection may actually improve the accuracy of a predictive model learnt on the smaller set of features.

- Reduces curse of dimensionality
- Helps with overfitting
- Reduces noisy features

