Quiz 2

① True. In model selection stage, the model with the smallest training error is chosen for final model assessment. Then, the model complexity parameters (hyperparameters if any) are determined by evaluating the model on validation sets. The model with the smallest training error implies it has fit the training data really well and if the test error of the model is also small, then, the model generalizes well on unseen data as well. Thus, training error and test error are important for finalizing a model.

②

Clustering is done by finding centers of data clusters such that the distance from the center from the data points in each cluster is minimum.

Since, vector qunatization based lossy compression of images requires forming centered-clusters of pixels, clustering can be used for vector quantization.

③

(a) When we apply non-linear functions to linear regression, it can be used for non-linear prediction from data.

For Example: In binary classification problem, we apply the sigmoid (non-linear) function to linear regression and then classify data according to whether they lie above or below the decision boundary.

(b) Regularization in linear regression optimization reduces the weights of the classifier. This makes sure that the weights (W) of the classifier are restricted, thereby, lowering the degree of freedom and flexibility of the classifier. This helps reduce overfitting of data.

(4)

$$A = \begin{bmatrix} 8 & -2 \\ -2 & 8 \end{bmatrix}$$

$$|A - \lambda I| = \begin{vmatrix} 8-\lambda & -2 \\ -2 & 8-\lambda \end{vmatrix} = 0$$

$$= (8-\lambda)(8-\lambda) - 4 = 0$$

$$\Rightarrow 64 - 8\lambda - 8\lambda + \lambda^2 - 4 = 0$$

$$\Rightarrow \lambda^2 - 16\lambda + 60 = 0.$$

$$\Rightarrow (\lambda - 6)(\lambda - 10) = 0$$

$$\boxed{\therefore \lambda_1 = 10 \quad \text{and} \quad \lambda_2 = 6}$$

Eigenvectors, v_1 and v_2,

$$\Rightarrow \begin{bmatrix} -2 & -2 \\ -2 & -2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 0$$

$$\Rightarrow u_1 = -u_2$$

$$\boxed{\therefore v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}}$$

and $$\Rightarrow \begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 0$$

$$\Rightarrow u_1 = u_2$$

$$\boxed{\therefore v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}}$$

(5)

(a) Left singular vectors of A,

$$u_1 = \frac{1}{2}\begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$u_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

(b) Right singular vectors of A,

$$v_1 = \frac{1}{2}\begin{bmatrix} 2 & 0 & 0 \end{bmatrix}$$

$$v_2 = \frac{1}{2}\begin{bmatrix} 0 & 0 & 2 \end{bmatrix}$$

(c) Singular values of A,

$$\sigma_1 = 1$$

$$\sigma_2 = 1$$

6

By using rank-k approximation, we get $k < \min(m,n)$.

Thus, the number of stored values will be,

$$\vartheta = (m+n+1)\cdot k < mn. \quad (\because k < \min(m,n))$$

Thus, a rank-k approximation of SVD can be used to make number of stored values less than mn.

7

Minimizing the cost function w.r.t W_k,

$$\frac{\partial J(k)}{\partial W_k} = -\frac{2}{N}\sum_{i=1}^{N}(x_i - W_k^T W_k x_i)\cdot 2W_k x_i$$

$$0 = -\frac{2}{N}\sum_{i=1}^{N}(x_i - W_k^T W_k \cdot x_i)\cdot 2W_k x_i$$

$$\Rightarrow \boxed{W_k^T W_k = 1.}$$

Thus, we chose k such that $W_k^T W_k = 1$, which implies we chose the value of k which minimizes the cost function $J(k)$.