Diep Vu

Homework 2

1 PCA can be explained from 2 different perspectives:

- **Maximum variance**: involves finding the directions of maximum variance in the data, which are also called the principal components
- **Minimum reconstruction error**: involves finding a low-dimensional representation of the data that minimizes the error between the original data and its reconstructed form

2 a. The scalar projection of x in the direction of w is $w_1^T x$

b. The scalar projection of $x - \mu$ in the direction w_1 is $w_1^T (x - \mu)$

c. The first component of y: $y_1 = w_1^T (x - \mu) / \sqrt{\lambda_1}$

d. $\tilde{x} = PCA^{-1}(y)$, y only has 1 component

$$\tilde{x} = y_1 \sqrt{\lambda_1}\, w_1 + \mu$$

e. $\tilde{x} = PCA^{-1}(y)$, x and y have the same number of elements

$$x - \tilde{x} = 0$$ (since $K = M$ or $x = y$)

3

4
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

a. NLL loss function:

$$\text{Loss} = -\log f_{X_1, \dots, X_N}(x_1, \dots, x_N) = -\sum_{n=1}^{N} \log f(x_n)$$

$$= -\log(\lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2} \cdots \lambda e^{-\lambda x_N})$$

$$= -\log(\lambda^N \cdot e^{-\lambda(x_1 + x_2 + \dots + x_N)})$$

b.
$$\text{Loss} = -\log \lambda^N - \log e^{-\lambda(x_1 + x_2 + \dots + x_N)}$$

$$= -N \log \lambda + \lambda(x_1 + x_2 + \dots + x_N)$$

$$\frac{d\,\text{Loss}}{d\lambda} = \frac{-N}{\lambda} + (x_1 + x_2 + \dots + x_N) = 0$$

$$\rightarrow \frac{N}{\lambda} = x_1 + x_2 + \dots + x_N \rightarrow \lambda = \frac{N}{x_1 + x_2 + \dots + x_N}$$