

Quiz 04

October 1st, 2026

15 minutes | 5 points

Name _____

Instructions. Answer both parts. Show your work and briefly justify each answer. Each part is worth 2.5 points. Use the space provided.

1. Error and retained variance (Homework 03, Problem 3)

5 points

Let $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^d$ be observations with mean $\bar{\mathbf{x}}$, and let $\mathbf{X} \in \mathbb{R}^{n \times d}$ be the centered data matrix with rows $(\mathbf{x}_i - \bar{\mathbf{x}})^\top$. Let $S \subseteq \mathbb{R}^d$ be a subspace with orthonormal basis $\mathbf{w}_1, \dots, \mathbf{w}_k$, write $\mathbf{W} = [\mathbf{w}_1 \cdots \mathbf{w}_k]$, and let $\mathbf{P}_S = \mathbf{W}\mathbf{W}^\top$ be the orthogonal projection onto S . You may use that $\mathbf{P}_S^\top = \mathbf{P}_S$ and $\mathbf{P}_S^2 = \mathbf{P}_S$.

Recall the *total variance*, the *variance retained by S* , and the *reconstruction error* of the data on S :

$$\text{TV} = \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i - \bar{\mathbf{x}}\|_2^2, \quad \text{TV}(S) = \frac{1}{n} \sum_{i=1}^n \|\mathbf{P}_S(\mathbf{x}_i - \bar{\mathbf{x}})\|_2^2,$$

$$\mathcal{E}(S) = \sum_{i=1}^n \|(\mathbf{x}_i - \bar{\mathbf{x}}) - \mathbf{P}_S(\mathbf{x}_i - \bar{\mathbf{x}})\|_2^2.$$

(a) Let $\mathbf{x} \in \mathbb{R}^d$. Show that $\mathbf{P}_S \mathbf{x}$ and $\mathbf{x} - \mathbf{P}_S \mathbf{x}$ are orthogonal.

Solution. Using $\mathbf{P}_S^\top = \mathbf{P}_S$ and then $\mathbf{P}_S^2 = \mathbf{P}_S$,

$$(\mathbf{P}_S \mathbf{x})^\top (\mathbf{x} - \mathbf{P}_S \mathbf{x}) = \mathbf{x}^\top \mathbf{P}_S \mathbf{x} - \mathbf{x}^\top \mathbf{P}_S^2 \mathbf{x} = \mathbf{x}^\top \mathbf{P}_S \mathbf{x} - \mathbf{x}^\top \mathbf{P}_S \mathbf{x} = 0,$$

so $\mathbf{P}_S \mathbf{x}$ and $\mathbf{x} - \mathbf{P}_S \mathbf{x}$ are orthogonal.

(b) Show that

$$\text{TV} = \text{TV}(S) + \frac{1}{n}\mathcal{E}(S).$$

Solution. For any $\mathbf{x} \in \mathbb{R}^d$, the vectors $\mathbf{P}_S \mathbf{x}$ and $\mathbf{x} - \mathbf{P}_S \mathbf{x}$ are orthogonal by part (a) and sum to $\mathbf{x}$, so the Pythagorean theorem gives

$$\|\mathbf{x}\|_2^2 = \|\mathbf{P}_S \mathbf{x}\|_2^2 + \|\mathbf{x} - \mathbf{P}_S \mathbf{x}\|_2^2.$$

Apply this with $\mathbf{x} = \mathbf{x}_i - \bar{\mathbf{x}}$ for each $i = 1, \dots, n$, sum over i , and divide by n :

$$\underbrace{\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i - \bar{\mathbf{x}}\|_2^2}_{\text{TV}} = \underbrace{\frac{1}{n} \sum_{i=1}^n \|\mathbf{P}_S(\mathbf{x}_i - \bar{\mathbf{x}})\|_2^2}_{\text{TV}(S)} + \underbrace{\frac{1}{n} \sum_{i=1}^n \|(\mathbf{x}_i - \bar{\mathbf{x}}) - \mathbf{P}_S(\mathbf{x}_i - \bar{\mathbf{x}})\|_2^2}_{\mathcal{E}(S)}.$$