

Machine Learning

V. Adamchik

CSCI 567

Spring 2019

Discussion 12

University of Southern California

# Principal Components Analysis

# Problem 1

---

Given 3 data points in 2-d space,  $(1, 1)$ ,  $(2, 2)$  and  $(3, 3)$ , what is the first principal component?

## Problem 2

---

The trace  $\text{Tr}$  of a square matrix is defined as the sum of the elements on the main diagonal. Prove that for the covariance matrix  $C$

$$\text{Tr}[C] = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2 = \lambda_1 + \lambda_2 + \dots + \lambda_n$$

where  $\lambda_i$  and  $\sigma_i$  are eigenvalues and variances.

# Problem 3

---

Find the second PCA value:

$$\begin{aligned} & \max(v_2^T X^T X v_2), \\ & \text{subject to} \\ & v_2^T v_2 = 1 \\ & v_2^T v_1 = 0 \end{aligned}$$

# Problem 4

---

Which of the following is wrong about PCA?

- (A) The first step of kernel PCA is to center the original dataset.
- (B) The first principal component is the eigenvector of the covariance matrix with the largest eigenvalue.
- (C) PCA outputs a compressed dataset that is a linear transformation of the original dataset.
- (D) Kernel PCA requires computing eigenvalues and eigenvectors of the Gram matrix.