EECS 551
Discussion 5

Zongyu Li

University of Michigan

October 1, 2021

Today's goals

1. Finish task-2-classify-subspace.ipynb.
2. Time-left questions.

Recall: Nearest neighbor classifier

- Pros:
  1. Intuitive, straight forward.
  2. No assumption on the data, unlike the LDA classifier.
- Cons:
  1. $\mathcal{O}(n)$ for testing which means it depends on the number of test examples.
  2. Outlier sensitivity.
  3. ...

Feature space

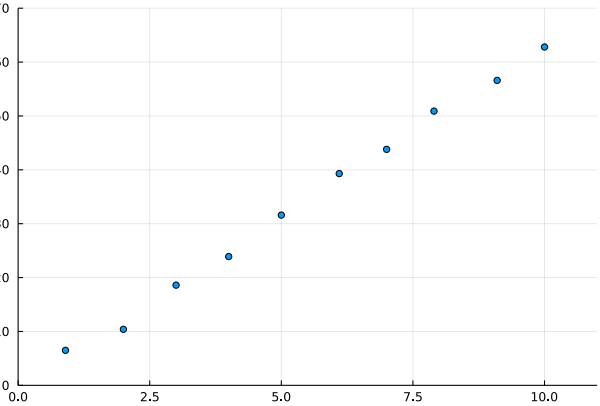
- Given data $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n] \in \mathbb{F}^{d \times n}$. Each $\boldsymbol{x}_i$ is a column vector of length d.
- Idea: Extract features from $\boldsymbol{X}$ using SVD.
- But how is the feature of $\boldsymbol{X}$ related to SVD?

Example: Suppose you don't know the relationship between radius and perimeter of circles. You took some measurements (noisy):

Radius (cm)	0.9	2.0	3.0	4.0	5.0	6.1	7.0	7.9	9.1	10.0
Perimeter (cm)	6.5	10.4	18.6	23.9	31.6	39.3	43.8	50.9	56.6	62.8

Feature space

- Scatter plot:



Feature space

- SVD:

$$\boldsymbol{X} = \begin{bmatrix} 0.9 & 2.0 & \dots & 10.0 \\ 6.5 & 10.4 & \dots & 62.8 \end{bmatrix} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}',$$

  where

$$\boldsymbol{U} = \begin{bmatrix} -0.157 & -0.988 \\ -0.988 & 0.157 \end{bmatrix},$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 125.1 & 0 \\ 0 & 0.5 \end{bmatrix}.$$

  One can see 125.1 is the dominating singular value.

- Question: How is $\boldsymbol{u}_1 = [-0.157 \ -0.988]'$ related to 2π?

Feature space

- From this example, we can see features of $\boldsymbol{X}$ are strongly related to left singular vectors that correspond to leading singular values of $\boldsymbol{X}$.
- Similarly, suppose that $\sigma_1, \sigma_2, ..., \sigma_k$ are the dominated singular values, then we can represent the feature space of $\boldsymbol{X}$ by $\boldsymbol{U}_k \triangleq [\boldsymbol{u}_1, \boldsymbol{u}_2, ..., \boldsymbol{u}_k]$.
- So the classifier can be designed by:
  1. Calculate $\boldsymbol{U}_k$ for each class using training data.
  2. For each test data, project onto the subspace formed by $\boldsymbol{U}_k$ for each class.
  3. Choose the nearest subspace and assign the corresponding label.

Review: project onto subspace

- **Todo:** Review lecture notes 3.62–3.67.
- Group work: Suppose that $\boldsymbol{P}$ is a projection matrix that maps vectors $\boldsymbol{x} \in \mathbb{F}^N$ onto a subspace $\boldsymbol{W}$.
  1. Prove or disprove: $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{F}^N$, $\|\boldsymbol{P}\boldsymbol{x} - \boldsymbol{P}\boldsymbol{y}\|_2 \leq \|\boldsymbol{x} - \boldsymbol{y}\|_2$.
  2. If $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{P}\boldsymbol{x}$, solve for $\boldsymbol{x}$ in terms of $\boldsymbol{y}$ and $\boldsymbol{P}\boldsymbol{y}$.

Task 2

Open the notebook
task-2-classify-subspace.ipynb.
(submit to gradescope using template.)