(a) Let the matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$ and $\mathbf{y} \in \mathbb{R}^{n}$ and no
**that $n > d$ and $\textbf{rank}(\mathbf{X} + \boldsymbol{\epsilon}_{\mathbf{X}}) = d$, explain**
**satisfying (5).**

If rank$(X + \epsilon_x)$ is $d$ = full rank,

therefore, for a fixed $y$, ~~there~~ exists a unique solution that makes $[X, y] \begin{bmatrix} w \\ -1 \end{bmatrix} = 0$, $w$ of that

, and $y$ are linearly dependent, its rank is deficient.

(b) Before continuing, we establish some linear algebra background.
trary matrix. Recall that the *Frobenius norm* of $\mathbf{A}$ is defined as $\|$
which can be thought of as the $\ell_2$-norm of $\mathbf{A}$ viewed as one long
$\|\mathbf{A}\|_F = \sqrt{\operatorname{tr}\left(\mathbf{A}^\top \mathbf{A}\right)}$**, and that if that $\mathbf{O} \in \mathbb{R}^{n \times n}$ and $\mathbf{P} \in \mathbb{R}^{m \times}$**
**then**

$$\|\mathbf{OAP}\|_F = \|\mathbf{A}\|_F,$$

$\| \cdot - U\Sigma V^T \|_F^2 = \mathrm{tr}\left( V\Sigma^T U^T \ldots \right) \sigma_i^2$

$\mathrm{tr}(\Sigma^T \Sigma) = \sum_{k=1} \sigma_i^2$

$U = U_{k \times d}$

$\Sigma = \Sigma_{k \times d}$

$V^T = V^T_{k \times d}$

(d) We will now leverage the low rank approximation to de
solution $\mathbf{w}$ to be unique, the matrix $[\mathbf{X} + \boldsymbol{\epsilon}_{\mathbf{X}}, \mathbf{y} + \boldsymbol{\epsilon}_{\mathbf{y}}]$ must
columns. Since this matrix has $d+1$ columns in total, it mus
the Eckart-Young-Mirsky Theorem tells us that the closest
norm is obtained by discarding the smallest singular val
$[\mathbf{X} + \boldsymbol{\epsilon}_{\mathbf{X}}, \mathbf{y} + \boldsymbol{\epsilon}_{\mathbf{y}}]$ that minimizes

$$||[\boldsymbol{\epsilon}_{\mathbf{X}}, \boldsymbol{\epsilon}_{\mathbf{y}}]||_F^2 = ||[\mathbf{X}^{true}, \mathbf{y}^{true}] -$$

is given by

$$[\mathbf{X} + \boldsymbol{\epsilon}_{\mathbf{X}}, \mathbf{y} + \boldsymbol{\epsilon}_{\mathbf{y}}] = \mathbf{U} \begin{bmatrix} \boldsymbol{\Sigma}_d & \\ & \end{bmatrix}$$

$$= \|(\epsilon_x, \epsilon_y)\|_F = \| - \begin{pmatrix} \cdot \\ U_{yy} \end{pmatrix} \begin{pmatrix} \sigma_{d+1} \end{pmatrix} \begin{pmatrix} \cdot \\ V_{yy} \end{pmatrix} \|_F$$

$$(\epsilon_x, \epsilon_y) = - \begin{pmatrix} U_{xy} \\ U_{yy} \end{pmatrix} \sigma_{d+1} \begin{pmatrix} V_{xy} \\ V_{yy} \end{pmatrix}^T$$

(e) **Using the result from the previous part and the fact t**

**Least Squares), find a nonzero solution to $[\mathbf{X} + \epsilon_{\mathbf{X}}, \mathbf{y}$**

**w in Equation (5).**

*HINT: Looking at the last column of the product $[\mathbf{X}, \mathbf{y}]$*
*problem, depending on how you solve it.*

As covered in the lecture note, such $\begin{pmatrix} w \\ -1 \end{pmatrix} =$

(a) **What is the $ij$-th entry of the matrices $\mathbf{X}\mathbf{X}^\top$ and $\mathbf{X}^\top\mathbf{X}$ in tern the matrix $\mathbf{X}\mathbf{X}^\top$ in terms of $\mathbf{U}$ and $\Sigma$, and, express the mat $\mathbf{V}$.**

(b) **Show that**

$$\psi_{\text{PCA}}(\mathbf{x}_i)^\top \psi_{\text{PCA}}(\mathbf{x}_j) = \mathbf{x}_i^\top \mathbf{V}_k \mathbf{V}_k^\top \mathbf{x}_j \quad \text{where} \quad \mathbf{V}_k =$$

Also **show that** $\mathbf{V}_k\mathbf{V}_k^\top = \mathbf{V}\mathbf{I}^k\mathbf{V}^\top$, where the matrix $\mathbf{I}^k$ denotes first $k$ diagonal entries as 1 and all other entries as zero.

(a) $ij^{th}$ elem $XX^T$ = $i^{th}$ row of $U$, scaled by $\Sigma^2$ then $j^{th}$ col of $U^T$

$\sum_{j=1}^{d} \sum_{i=1}^{d} u_i \Sigma^2 (u_j)^T$

$ij^{th}$ elem $X^TX$ = same as above but $V$ and

---

(b) $\phi PCA(x_i) = (V^T x_i)$ $\quad \phi PCA(x_i)^T = ($

$\phi PCA(x_j) = (V^T x_j)$ $\quad = \quad x$

$VV^T = V\underline{I}V^T$

$V^{-1}VV^T = IV^T$

Screen clipping taken: 9/30/2018 7:50 PM

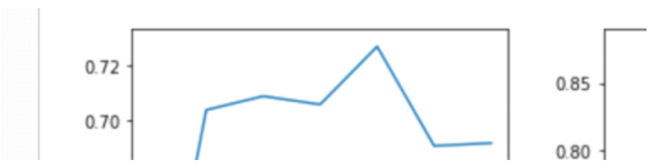
```python
## Input: original dimension d,
## Input: embedding dimension k
## Output: d x k random
## Gaussian matrix J with entry-w
## variances 1/k so that,
## for any row vector z^T in R^d,
## z^T J  is a random features em
def random_JL_matrix(d, k):
    return np.random.normal(loc=0


## Input: n x d data matrix X
## Input: embedding dimension k
## Output: d x k matrix V
## with orthonormal columns
## corresponding to the top k rig
## of X. Thus, for a row vector z
## z^T V  is the projection of z^
## onto the the top k right-singu
def pca_embedding_matrix(X, k):
    u, s, v = np.linalg.svd(X, 0)

    return v.T[:k].T
```

Screen clipping taken: 9/30/2018 8:00 PM

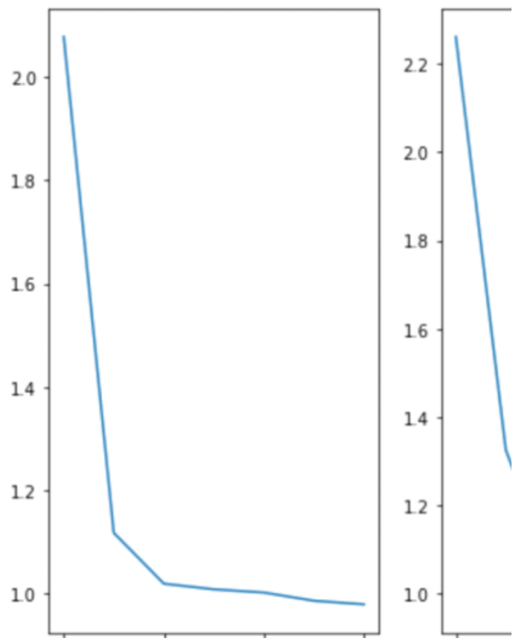
(f) For each dataset, we will now fit a linear model on differe
classification. The code for fitting a linear model with pro
for a given feature, is given to you. Use these functions a
in the following way: (1) Use top $k$-PCA features to ob
$k$-dimensional random embeddings to obtain the second
over 10 random embeddings for smooth curves). Use the



the starter code to select these features. You should vary $k$
of each feature $\mathbf{x}_i$. **Plot the accuracy for PCA and Rand**
**Comment on the observations on these accuracies as a**
**datasets.** Attach your plots below.



(g) Now plot the singular values for the feature matri
below. **Do you observe a pattern across the thr
performance of PCA observed in the previous p**




Screen clipping taken: 9/30/2018 8:05 PM

CS 189 Introduction to Machine
Fall 2018

## 1 Getting Started

**Read through this page carefully.** You may type
handwritten/scanned solutions. Please start each que

1. Submit a PDF of your writeup to assignment c
graphs, include those graphs in the correct sect

(a) Who else did you work with on this homework
group. How did you work on this homework? A

Just myself

(b) Please copy the following statement and sign ne
that no one inadverdently cheats.

*I certify that all solutions are entirely in my v*
*student's solutions. I have credited all external s*