CS 189 Introduction to Machine Learning
Fall 2018

HW4 Common Errors

On this homework, the most common reason for deducted points was insufficient justification. Keep in mind that many of the homework problems state the final expression, and therefore the bulk of the work is in showing that this expression is true, not in discovering its form. As such, I expect every step to be justified to show me that you understood your manipulations, and were not just trying to fit things into the form of the given expression.

Common Problem 2 Errors

(a)
- It was necessary to explicitly use the assumption on $rk(\mathbf{X})$ to argue for a rank of at least d.

(b)
- For the Frobenius norm property, it was necessary to show an intermediate step considering the form of the matrix multiplication $\mathbf{A}^\top \mathbf{A}$.
- The matrix multiplication rule for trace only allows for arguments to be reordered *cyclically*.

(c)
- I did not penalize for this, but the result in part (b) was shown for square orthonormal matrices, and thus some care should be taken when applying it to the compact SVD, where $\mathbf{U}$ is not square.

(d)
- I was looking for some justification of the following equality, either in words/sketches or through an intermediate matrix multiplication step:

$$\begin{bmatrix} \mathbf{U}_{xx} & \mathbf{u}_{xy} \\ \mathbf{u}_{yx}^\top & u_{yy} \end{bmatrix} \begin{bmatrix} 0 & \\ & \sigma_{d+1} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{xx} & \mathbf{v}_{xy} \\ \mathbf{v}_{yx}^\top & v_{yy} \end{bmatrix}^\top = \begin{bmatrix} \mathbf{u}_{xy} \\ u_{yy} \end{bmatrix} \sigma_{d+1} \begin{bmatrix} \mathbf{v}_{xy} \\ v_{yy} \end{bmatrix}^\top$$

Common Problem 4 Errors

(a)
- For full credit, it was necessary to show at least one step between the initial matrix expression and the final expression.
- The problem asked for the expression in terms of elements of the matrix, i.e. x_{ij}, so an expression in terms of its rows or columns was incomplete.

(b)
-

(c)
- For full credit, it was necessary to justify why $\sum_{i,j} \psi(\mathbf{x}_i)^\top \psi(\mathbf{x}_j) = \|\mathbf{X}\mathbf{V}_k\mathbf{V}_k^\top\mathbf{X}^\top\|_F$ by pointing out the expression of the ijth element or defining and reasoning about a matrix $\mathbf{\Psi}$. (The expression $x_i^\top \mathbf{V}_k \mathbf{V}_k^\top x_j$ was not sufficient on its own).
- It is not true that $\|\mathbf{A} - \mathbf{B}\|_F = \|\mathbf{A}\|_F - \|\mathbf{B}\|_F$.

(d)
- `np.random.normal` and similar methods take the *standard deviation* rather than the variance as arguments.
- `np.linalg.svd` returns `Vh`$= \mathbf{V}^\top$, and therefore it was necessary to take a transpose.

(e)
- The question asked about the differences between PCA and random embeddings and the trend over the 3 datasets, so both should have been addressed in your comments.
- I was fairly generous with credit for plot labelling, but to get full credit, it was necessary to color the scatter plot by y label, indicate if a plot was PCA or random, and indicate which dataset it came from.

(f)
- Similar errors to (e).

(g)
- The question asked about the differences between singular values over the 3 datasets and for a connection to the previous questions, so both should have been addressed in your comments.
- The singular value plots should have been clearly labeled with dataset number.