CS 189 Introduction to Machine Learning
Fall 2018

# HW1 Common Errors

## Common Problem 3 Errors

(a)
- Sloppy attempts at conditioning, i.e. "if $X \leq t$, $P(X \geq t) = 0$".

(b)
- Skipping too much of the integral or saying you used Wolfram Alpha. When doing the integral is most of the problem, you should write out some steps.
- Not referencing symmetry or something similar to invoke $P(|Z| \geq t) = 2P(Z \geq t)$.

(c)
- Missing the $t \geq 1$ step.
- Claiming the bound holds with equality.

(d)
- Using the two-sided bound without noting that $P(Z \geq t) \leq P(|Z| \geq t)$.
- Using the one-sided bound (part (b)) but claiming $t \geq 1$ (or a similar dropping of terms) without proof. You should instead verify it based on the possible range of $\delta$. Equivalently, making an unnecessary assumption on the size of $\theta_{\max} - \theta_{\min}$ or $T$.

(e)
- Similar errors as in (d).

## Common Errors in Problem 5

## Problem 5a

1. Problem 5a required that you prove an if-and-only-if. Points would be deducted if one direction was missing.

2. In Problem 5a, as in the rest of problem 5, we gave you the definition of a rank-$d$ projection matrix a symmetric matrix $P$ for which $P = P^\top$ and $P^2 = P$ and has rank $d$. Many solutions projected that this was in fact a projection operator, but this was not stated in the definition given (this point was also clarified in a Piazza post). If you assume that $P$ indeed corresponded to a projection operator without justification, points were deducted.

3. Many student solutions used the spectral decomposition, which required that you compute the eigenvalues of $P$. Points where deducted if students did not show that the eigenvalues of $P$ where in $\{0, 1\}$, or if the students did not account for the zero eigenvalues of $P$ in a meaningful way. For example, if $P = e_1 e_1^\top$ where $e_1 = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \end{bmatrix}$, then $P$ has $n-1$ zero-eigenvalues. If $P$ has no zero-eigevalues, then $\text{rank}(P) = 1$, and one can show that this necessarily implies $P$ is the identity, because you aren't "projecting away" any components.

4. Many students confused the singular value and spectral decomposition. If your solution tried to relate the singular and spectral decompositions, or the singular values and eigenvlaues *without adequate justification*, points were deducted. There were in fact valid solutions that used both the spectral and the signular value decomposition. Moreover, it turns at that for any projection $P$, the spectral and

signular value decompositions are actually the same. This is because $P$ is symmetric and postiive semidefinite. However, justifications of these points was needed for full credit.

5. Many students were unclear with their dimensions, especially when using the spectral decomposition. The cannonical spectral decomposition expresses $P = V\Lambda V^\top$, where $V, \Lambda \in \mathbb{R}^{n\times n}$. However, the solution asked for a $U \in \mathbb{R}^{n\times d}$. While subtle, being clear with dimensions allows the grader to understand that you realize $P$ has zero eigenvalues, and that they must be accounted for. Alternatively, you could have uses the rank of $P$ to take a compact spectral decomposition, but you would have had to make this clear from the outset to receive full credit. Many students also lost points for being unclear with the dimensions when applying the QR decomposition, though this is in general a perfectly acceptable way to solve the problem when dimensions are accounted for.

6. To receive full credit, you needed to check that $U^\top U = I$. If you derived $U$ from a spectral decomposition or SVD, you got this for free. Many students tried to use the fact that $UU^\top UU^\top = UU^\top$ to implie that $U^\top U = I$. THis is not immediately clear, because $U \in \mathbb{R}^{n\times d}$ can't be naively inverted (it is rectangular). To be rigorous, you would have needed to use a pseudoinverse, or provided an alternative explanation.

7. Many students assumed that $P = UU^\top$, and then showed that $U^\top U$ must be the identity. This line of argumentation often had other flaws, but the most glaring is that it requires that you justify why you can write $P = UU^\top$ for some $U$. Students also tried writing $P = UV^\top$, and arguing that necesarily $V = U$. This is flat out wrong, since I could always write $P = IP^\top$, and $I \neq P$.

# Problem 5b

Most students did quite well on this problem. Here were some common mistakes:

1. Again, we found some students confusing the SVD and spectral decomposition.

2. Some students were not clear in distinguishing between $P$ and its eigenvalue-diagonal matrix $D$; some solutions seemed to claim that $P$ was diagonal. THis is in general not true. For example, if $v = \mathbf{1}/\sqrt{n}$, where $\mathbf{1}$ is the one's vector in $\mathbb{R}^n$, then $P = vv^\top$ is a projection matrix, but is not a diagonal.

3. Many students correctly attempted to solve the problem by relating eigenvalues to trace. However, they did not justify that the eigenvalues were in $\{0, 1\}$, and therefore did not complete the necessary computation. More generally, some students did not justify other parts of their trace computation.

4. Many students tried to take $U^{-1}$, where $U \in \mathbb{R}^{n\times d}$ was the matrix in the decomposition from Part $a$. This is simply incorrect, because you cannot invert a rectangular matrix.

# Problem 5c

1. By far the most common error was assuming $P$ to be the matrix which performs an orthogonal projection, rather than establishing this fact from the properties provided; i.e. that $P = P^\top$, and $P^2 = P$. We clarified this point on Piazza. Students who correctly proved that $P$ was an orthogonal projection recieved full credit.

2. Many students tried to direction show $Pv$ mimized $\|w - Pv\|_2^2$ by expanding

$$\|w - Pv\|_2^2 = \|w - v\|_2^2 + \|v - Pv\|_2^2 + 2\langle w - v, v - Pv\rangle$$

 

This is a totaly valid plan of attack, but many students (a) neglected to show the cross term (or justify why it was not there), (b) would write out the cross term but set it to zero and not justify it, or (c) using the fact that $P$ was a projection (which they had not yet established).

3. A couple students tried to argue that $Px = Pv$ implies $x = v$, which is in general not true. Other students tried to inver $P$, which is not possible when $P$ is rank deficient.

4. Many students tried computing gradients of $\|v - w\|_2^2$ with respect to $w$. There is a way to do this correctly (using, e.g. Lagrangians or KKT conditions), and you may use these if you are comfortable with them. However, many students argued in incorrect or unrigorous ways, and consequently lost points.

5. Many students seemed to present nearly identitcal solutions by expanding $\|v - P(v + \Delta)\|_2^2$, and concluding $\|P\Delta\|_2^2 = 0$. I took off one point, because it seemed that like students didn't provide a clear concluding line, like "If $w = P(v + \Delta)$ is a minimizer, then $P\Delta = 0$, so that $w = Pv$". It is imortant to explain your reasoning, and since so many students presented the *exact* same soutions, I wanted to see proper explanation. Strings of equations without explanation in general may not receive full credit. Second, *never* use $d$ as a vector, especially when it shows up as a problem parameter elsewhere.

# Problem 5d

1. Many students did not seem to read the erata, which clearly stated that $X \in \mathbb{R}^{n\times d}$ (rather than $X \in \mathbb{R}^{d\times d}$ as stated on the homework). Given that the erata was posted, pinned to the Piazza site, and an email was sent out, we held students responsible for being aware of this change. Moreover, if $X \in \mathbb{R}^{d\times d}$, the problem because absolutely trivial, which should have alerted you to an error. Solutions which all required $X^{-1}$ to exists lost all credit; solutions which used $X^{-1}$ is certain parts of the solutions but not others lost partial credit, depedning on how strongly they relied on $X^{-1}$.

2. Students lost points for not verifying not verifying that $X(X^TX)^{-1}X^T$ had rank $d$.

3. Students lost points for not getting the correct dimension of the $U$ matrix, or for providing a $U$ matrix and not verifying that it satisfies $U^\top U = I_{d\times d}$. Students who used a QR-decompsition often has issues specifying the dimensions correctly

4. Students did not provide a transparent interpretation of $U$; either then said it was just the matrix from part (a), or they expressed $U$ in terms of the SVD but without simplifying. Correct interpretations/answers for $U$ inlcuded: $U$ matrix from the SVD, eigenvectors from $X^\top X$, matrix whose columns are an orthonormal basis from the range/columnspace of $X$.

# Problem 5e

1. One point was deduced for not justifying the computation $\mathbb{E}[z^\top(\ldots)z] = \mathrm{tr}(\ldots)$. Acceptable ways to justify the computation was to show that

$$\mathbb{E}[z^\top Az] = \mathbb{E}[\sum_{i,j} A_{i,j}z_{i,j}] = \sum_{i,j} A_{i,j}\mathbb{E}[z_{i,j}] = \sum_{i,j} A_{i,j}\mathbf{I}(i=j) = \mathrm{tr}(A)$$

and

$$\mathbb{E}[z^\top Az] = \mathbb{E}[\mathrm{tr}(zz^\top A)] = \mathrm{tr}\Big(\mathbb{E}(zz^T)A\Big) = \mathrm{tr}(IA) = \mathrm{tr}(A)$$

Some people cited Wikipedia; this also recieved credit because at least it was a justification.

 

2. Many students dropped expectations in crucial places. For example, students may have claimed that $z^\top A z = \mathrm{tr}(A)$, which is false with probability 1. However, when expectations are taken, the equality holds.

3. Many students botched the trace cyclic rule, and got things like $\mathbb{E}[z^\top z \, \mathrm{tr}(A)]]$. Students lost a lot of points for this, because it actually yielded incorrect answers. For example, if you ended up with $\mathbb{E}[z^\top z] \, \mathrm{tr}\Big((X^\top X)^{-1}\Big)$, then $\mathbb{E}[z^\top z] = n$, and so the answer is off by a factor of $n$. In some cases, students would make multiple errors which would cancel eachother out. This resulted in more deductions :-).

4. Some students claimed tht $(X^\top X)^{-1}$ was a projection matrix. This is false.

5. $X$ is rectangular, and therefore not invertible.

# Problem 5f

1. Most students lost points for omitting the $\mathrm{rank}(X) < d$ case, giving a partial answer for the $\mathrm{rank}(X) < d$ case (e.g. an example when $\mathrm{rank}(X) = 0$, or saying error decreases), or correctly stating the error of $\mathrm{rank}(X)/n$, but not providing justification.

2. Some students said that if $\mathrm{rank}(X) < d$, error increases. This is just wrong.

3. When $\mathrm{rank}(X) < d$, some students tried to write $X(X^\top X)^{-1}X^\top$. This is ill defined, since $(X^\top X)$ is rank deficient.

4. Some students correctly used pseudoinverses to handle the low rank case, but did not justify the trace computation involving the pseudoinverse. One point was deducted for this. If the justification was false, two points were removd.

5. Some students claimed that $(X^\top X)^\dagger X^\top X$ was diagonal, which is false.