CS 189 HW5: comments and common errors

December 10, 2018

Problem 3a.

- significant algebra error in computing key quantities
- Slight errors: In disproving that the loss function is convex, showed less (more) than or equal, instead of *strictly* less (more) than. Does not justify direction of inequality correctly (by noting that λ_1 is strictly positive) .
- Proof attempt by non-rigorous/incomplete argument: e.g. did not compute f(0). e.g. did not check or justify that function value is *strictly* greater on the line between the 2 optimal points (assumed that there are *only* 2 optimal solutions. The problem does not ask you to assume this; it only said that $\pm\sqrt{\lambda_1}u_1$ are minima.)
- Proof attempt by a wrong claim. e.g. saying convex functions must have unique optimum (untrue unless strictly convex!) e.g. implying that f=0 is the minimal function value (False!)
- computed Hessian but did not justify when it isn't PSD.
- Didn't use correct definition of convex or didn't refer to definition of convexity at all or did not apply definition of convex correctly.

Problem 3b.

- Did not differentiate half of the expression correctly (got either $\|w\|^4$ or $w^T M w$ wrong) OR other single significant mistake e.g. created an extra trace term.
- tried to add a scalar with a matrix.
- sign error/extra scalar factor

Problem 3c.

The most common mistake was showing that given solution set is correct, BUT failing to prove that there are no other solutions. For this you must use the distinctness of the non-zero eigenvalues.

Also note that the solution set is the first r eigenvectors instead of d. I did not take off points if you made this typo.

Another common mistake is saying $ww^T = M$. This is FALSE! For A a matrix and v a vector, $Av = 0$ DOES NOT imply $A = 0$ or v. It does imply v is in the nullspace of A. Really should know this by now... :)

Problem 3d. (Not graded)
Self-check for mistakes:

- Hessian computed wrongly (correct Hessian is $\|w\|^2 I + 2ww^T - M$).
- Computed Hessian correctly but didn't find positive and/or negative eigenvalues of Hessian.

Problem 3e. (Not graded)
Self-check for mistakes:

- did not justify induction step fully (e.g. need to say that span$\{u_i\}_{i \in I}$ is an invariant subspace for M.)

Problem 3f.
Acceptable answer: uniformly random on a sphere, isotropic gaussian, etc.

- Wrong answer: any answer that requires knowledge of the eigenvectors of M (e.g. u_1, span$\{u_i\}_{i \in I}$ etc.) If you knew that that, you wouldn't need to run GD to find the top eigenvector.
- -1: Said random initialization without specifying what kind of randomness/specifying a distribution (or suggested a random scheme that has high prob of failing, e.g. choose a random column of M), but had some reasonable justification. It's simply not possible to "choose a random vector in $\mathbb{R}^d$" (a uniform distribution over the real line doesn't exist).
- -2: just said random initialization but answer doesn't add anything meaningful on top of the hint/lacks justification/incorrect justification (correct justification: want initialization to not be orthogonal to u_1. Note that this is NOT the same as "ant initialization to be in the span of u_1".)
- unnecessary comments about step size. The question is about initialization, step size is irrelevant here.

Problem 4b.
Many people wrote

$$\frac{\partial MSE}{\partial \hat{y}} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i). \tag{1}$$

This is incorrect because it is summing up the derivatives of functions $\hat{y} \to (\hat{y} - y_i)^2$ evaluated at different points. This is not a meaningful quantity.

The correct answer required treating each $\hat{y}_i$ as a sperate variable, a coordinate of $\hat{y}$. Therefore, $\frac{\partial MSE}{\partial \hat{y}}$ is a vector with coordinates $(\hat{y}_i - y_i)$. We also accepted an answer which look at a loss for only one data point and treated $\hat{y}$ as a scalar.

Problem 4c.

Many people wrote

$$\frac{\partial MSE}{\partial a_i} = \frac{\partial MSE}{\partial a_{i+1}} \sigma'(W_i a_i + b_i) W_i,$$

including the person who wrote the solutions last semester. This is incorrect because $\sigma'(W_i a_i + b_i)$ means that we apply the derivative of the activation σ coordinate-wise to the vector $W_i a_i + b_i$. Therefore, $\sigma'(W_i a_i + b_i)$ is a vector. However, in this case, we would get that $\frac{\partial MSE}{\partial a_i}$ is a matrix, which is wrong.

The correct answer is

$$\begin{aligned}\frac{\partial MSE}{\partial a_i} &= \frac{\partial MSE}{\partial a_{i+1}} \frac{\partial \sigma(z_i)}{z_i} W_i \\ &= \frac{\partial MSE}{\partial a_{i+1}} \operatorname{diag}(\sigma'(W_i a_i + b_i)) W_i,\end{aligned}$$

Note that $\sigma(z_i)$ is a vector so $\frac{\partial \sigma(z_i)}{z_i}$ is a matrix. The solution has been updated on bcourses.