Alireza Mohammadshafie

1. When does the least squares solution have a unique solution? (0.5 pt)

When our features are independent of each other (not linearly dependent), the least squares solution is unique. For example, if we have feature for height in cm, we shouldn't have another feature for height in inch or foot!

2. What is RANSAC and why is it a generic framework for robust fitting? (1 pt)

RANSAC is an algorithm that randomly picks points, fits a model, then checks which points are close. It's generic because it ignores outliers and can fit lots of different types of models in the same way.

3. Please describe the influence of learning rate on the convergence of gradient descent. (0.5 pt)

If the learning rate is too big, you jump around and never settle. If it's too small, you move super slowly and might take forever to finish. We have to find a decent learning rate with try and error usually.

4. Should we use the least squares for binary classification? Please also explain the reason. (2 pts)

No, because least squares is for continuous outputs and can give weird probabilities (like negative numbers). Binary classification needs something like logistic regression or any other classification technique like SVM to deal with yes/no or 0/1 properly.

5. Why could L1 regularization be used for feature selection? (1 pt)

Because it pushes some coefficients all the way to zero, effectively kicking out less important features and simplifying the model.