

Quiz 06

October 15th, 2026

15 minutes | 5 points

Name _____

Instructions. Answer all three parts. Show your work and briefly justify each answer. Parts (a) and (b) are worth 2 points each and part (c) is worth 1 point. Use the space provided.

1. Ridge regression with many features (Homework 05, Problem 1)

5 points

Let $\mathbf{X} \in \mathbb{R}^{n \times d}$ have rows $\mathbf{x}_1^\top, \dots, \mathbf{x}_n^\top$, let $\mathbf{y} \in \mathbb{R}^n$, and let $\lambda > 0$. Recall that $\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I}_d$ is invertible and that the ridge regression estimator is

$$\hat{\boldsymbol{\theta}}_\lambda = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^d} \left\{ \|\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\|_2^2 + \lambda \|\boldsymbol{\theta}\|_2^2 \right\} = (\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I}_d)^{-1} \mathbf{X}^\top \mathbf{y}.$$

This requires solving a $d \times d$ linear system. In this problem you will find a second formula that's instead an $n \times n$ system.

(a) Show that $\mathbf{X}\mathbf{X}^\top + \lambda \mathbf{I}_n$ is invertible. (2 points)

Solution. Since $\mathbf{X}\mathbf{X}^\top + \lambda \mathbf{I}_n$ is square, it suffices to show its null space is trivial. Suppose $(\mathbf{X}\mathbf{X}^\top + \lambda \mathbf{I}_n)\mathbf{z} = \mathbf{0}$. Multiplying on the left by $\mathbf{z}^\top$ gives

$$0 = \mathbf{z}^\top \mathbf{X}\mathbf{X}^\top \mathbf{z} + \lambda \mathbf{z}^\top \mathbf{z} = \|\mathbf{X}^\top \mathbf{z}\|_2^2 + \lambda \|\mathbf{z}\|_2^2 \geq \lambda \|\mathbf{z}\|_2^2.$$

Since $\lambda > 0$, this forces $\|\mathbf{z}\|_2 = 0$, so $\mathbf{z} = \mathbf{0}$.

- (b) Show that $(\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I}_d) \mathbf{X}^\top = \mathbf{X}^\top (\mathbf{X} \mathbf{X}^\top + \lambda \mathbf{I}_n)$, and conclude that (2 points)

$$\hat{\boldsymbol{\theta}}_\lambda = \mathbf{X}^\top (\mathbf{X} \mathbf{X}^\top + \lambda \mathbf{I}_n)^{-1} \mathbf{y}.$$

Solution. Expanding, both sides equal $\mathbf{X}^\top \mathbf{X} \mathbf{X}^\top + \lambda \mathbf{X}^\top$. Multiply the identity on the left by $(\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I}_d)^{-1}$, which exists as recalled above, and on the right by $(\mathbf{X} \mathbf{X}^\top + \lambda \mathbf{I}_n)^{-1}$, which exists by part (a):

$$\mathbf{X}^\top (\mathbf{X} \mathbf{X}^\top + \lambda \mathbf{I}_n)^{-1} = (\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I}_d)^{-1} \mathbf{X}^\top.$$

Applying both sides to $\mathbf{y}$,

$$\mathbf{X}^\top (\mathbf{X} \mathbf{X}^\top + \lambda \mathbf{I}_n)^{-1} \mathbf{y} = (\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I}_d)^{-1} \mathbf{X}^\top \mathbf{y} = \hat{\boldsymbol{\theta}}_\lambda.$$

- (c) Prove that $\hat{\boldsymbol{\theta}}_\lambda$ is a linear combination of the observations $\mathbf{x}_1, \dots, \mathbf{x}_n$. *Hint:* First, show that there is a vector $\boldsymbol{\alpha} \in \mathbb{R}^n$ with $\hat{\boldsymbol{\theta}}_\lambda = \mathbf{X}^\top \boldsymbol{\alpha}$. (1 point)

Solution. By part (b), take $\boldsymbol{\alpha} = (\mathbf{X} \mathbf{X}^\top + \lambda \mathbf{I}_n)^{-1} \mathbf{y} \in \mathbb{R}^n$. The columns of $\mathbf{X}^\top$ are $\mathbf{x}_1, \dots, \mathbf{x}_n$, so

$$\hat{\boldsymbol{\theta}}_\lambda = \mathbf{X}^\top \boldsymbol{\alpha} = \sum_{i=1}^n \alpha_i \mathbf{x}_i.$$