

Midterm 1: Math 6266

Peter Williams

Section 1.1

Exercise 1. Consider the linear regression model with mean zero, uncorrelated, heteroscedastic noise:

$$Y_i = X_i^\top \theta + \varepsilon_i, \text{ for } i = 1, \dots, n, \quad E\varepsilon_i = 0, \quad \text{cov}(\varepsilon_j, \varepsilon_j) = \begin{cases} \sigma_i^2, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \quad (1)$$

Find expressions for the *LSE* and response estimator in this model:

Under heteroscedastic noise assumptions, the *LSE* estimator is the same as the $\hat{\theta}_{OLS}$, which is:

$$\hat{\theta}_{OLS} = \underset{\theta}{\operatorname{argmin}} \|Y - X^\top \theta\|^2 = \underset{\theta}{\operatorname{argmin}} G(\theta)$$

,

$$\|Y - X^\top \theta\|^2 = G(\theta) = (Y - X^\top \theta)^\top (Y - X^\top \theta) = YY^\top - 2\theta^\top XY + \theta^\top XX^\top Y$$

with gradient,

$$\nabla G(\theta) = -2XY + 2\theta^\top XX^\top Y$$

Setting this expression equal to zero leads to estimator $\hat{\theta} = \hat{\theta}_{OLS} = (XX^\top)^{-1}XY$, which leads to response estimator $\hat{Y} = X^\top \hat{\theta} = X^\top (XX^\top)^{-1}XY$.

Exercise 2. Assume that $\varepsilon_i \sim N(0, \sigma_i^2)$ in the previous problem. What is known about the distribution of $\hat{\theta}$ and \hat{Y} ? Now suppose additionally that $\sigma_i^2 \equiv \sigma^2 > 0$. What can be said about distribution of the estimator $\hat{\sigma}^2$?

Exercise 3. Consider the linear regression model from exercise 1. Suppose, that the target of estimation is $h^\top \theta$ for some determinate non-zero vector $h \in R^p$. Find expression for the LSE of $h^\top \theta$. Is this estimate optimal in sense of Gauss-Markov theorem, i.e. does it have the smallest variance among all linear unbiased estimators? Study the same issue for the target $\eta = H^\top \theta$, where $H \in R^{q \times p}$ is some non-zero matrix with $q \leq p$.

Section 1.3

Exercise 4. Let $A \in R^{n \times n}$ be a matrix (corresponding to a linear map in R^n). Show that A preserves length for all $x \in R^n$ iff it preserves the inner product. I.e. one needs to show the following: $\|Ax\| = \|x\| \quad \forall x \in R^n \iff (Ax)^\top (Ay) = x^\top y \quad \forall x, y \in R^n$.

$$\begin{aligned} \|x\| = \sqrt{x \cdot x} = \sqrt{x^\top x} &\implies \|Ax\| = \sqrt{Ax \cdot Ax} = \sqrt{x^\top A^\top A x} \implies \\ A^\top A &= I_n = A^{-1}, \quad A^\top = A^{-1}, \quad \|Ax\| = \|x\| \end{aligned}$$

this implies A is an orthogonal matrix, and further,

$$(Ax)^\top (Ay) = \|Ax Ay\|^2 = x^\top A^\top A y = x^\top y = \|xy\|^2$$

Exercise 5. (a) Let $x_0 \in R^n$ be some fixed vector, find a projection map on the subspace $\operatorname{span}(x_0)$. Compare your result with matrix Π (from section 1.3) for the case of $p = 1$. (b) Prove part 3) of Lemma 1.1 for an arbitrary orthogonal projection in R^n . Exercise 6. Let $L1, L2$ be some subspaces in R^n , and $L2 \subseteq L1 \subseteq R^n$. Let $PL1, PL2$ denote orthogonal projections on these subspaces. Prove the following properties: (a) $PL2 - PL1$ is an orthogonal projection, (b) $|PL2| \leq |PL1| \quad \forall x \in R^n$, (c) $PL2 \cdot PL1 = PL2$

Section 2.1

Exercise 7. (a) Using the notation from section 2.1, consider $X \sim N(\mu, I_n)$ for some $\mu \in R^n$. Find $EQ(X)$ and $VarQ(X)$. (b) Generalize the results from part (a) to the case $X \sim N(\mu, \Sigma)$ for some positive-definite covariance matrix $\Sigma \in R^{n \times n}$.

Exercise 8. Let $X \sim N(0, I_n)$, $Q = XX^T$. Suppose that Q is decomposed into the sum of two quadratic forms: $Q = Q_1 + Q_2$, where $Q_i = X^T A_i X$, $i = 1, 2$ for some symmetric matrices A_1, A_2 with $rank(A_1) = n_1$ and $rank(A_2) = n_2$. Show that if $n_1 + n_2 = n$, then Q_1 and Q_2 are independent and $Q_i \sim \chi^2(n_i)$ for $i = 1, 2$.

Section 2.2

Exercise 9. In the Gaussian linear regression model 3, consider the target of estimation $\eta = H^T \theta^*$, where $H \in R^{q \times p}$ is some non-zero matrix with $q \leq p$. Find an analogue of the quadratic form S_2 (from (4)) for the new target η^* , and prove for the new quadratic form statements similar to (e) from Theorem 2.1, and Corollary 2.1.2.

Exercise 10. (a) Consider model (3) for $p = 2$, $X_i = (1, x_i)^T$, $\theta^* = (\theta_1^*, \theta_2^*)^T$ (similarly to section 1.5). Write explicit expressions for the confidence sets for $\theta^*, \theta_1^*, \theta_2^*$.

(b) Find a confidence interval for the expected response $E[Y_i]$ in the model in part (a).

Exercise 11. Find an elliptical confidence set for the expected response $E[Y]$ in model (3).

Exercise 12. Construct simultaneous confidence intervals (e.g., as in Corollary 2.2.1) for the expected responses $E[Y_1], \dots, E[Y_n]$ in model (3).