Midterm 1: Math 6266

Peter Williams

Section 1.1

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

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

Find expressions for the LSE and response estimator in this model

Under heteroscedastic noise asssumptions, the LSE estimator, denoted $\hat{\theta}_{OLS}$, is:

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

,

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

with gradient,

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

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

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 $\hat{Y}$?

Denote $n \times n$ matrix $D = diag\{\sigma_1^2, \sigma_2^2, ..., \sigma_n^2\} = Var(\varepsilon)$.

For $\hat{\theta}$, we have,

$$E[\hat{\theta}] = E[(XX^\intercal)^{-1}XY] = E[(XX^\intercal)^{-1}X(X^\intercal \theta^* + \varepsilon)] = E[\theta^*] + E[\varepsilon] = \theta^*$$

indicating that $\hat{\theta}$ is unbiased despite the presence of heteroscedastic noise. Further $\hat{\theta}$ is normally distributed, since is a linear transformation of $\varepsilon \sim N(0, D)$. Further we have,

$$Var(\hat{\theta}) = Var((XX^\intercal)^{-1}XY) = Var((XX^\intercal)^{-1}X(X^\intercal \theta^* + \varepsilon)) = Var((XX^\intercal)^{-1}X\varepsilon)) =$$
$$(XX^\intercal)^{-1}XVar(\varepsilon)X^\intercal (XX^\intercal)^{-1} = (XX^\intercal)^{-1}XDX^\intercal (XX^\intercal)^{-1} = Var(\hat{\theta})$$

For $\hat{Y}$ we have,

$$E[\hat{Y}] = E[X^\intercal (XX^\intercal)^{-1}XY] = E[X^\intercal (XX^\intercal)^{-1}X(X^\intercal \theta^* + \varepsilon)] = E[X^\intercal \theta^* + X^\intercal (XX^\intercal)^{-1}X\varepsilon] = E[X^\intercal \theta^*] = Y$$

and,

$$Var[\hat{Y}] = Var[X^\intercal (XX^\intercal)^{-1}XY] = Var[X^\intercal (XX^\intercal)^{-1}X(X^\intercal \theta^* + \varepsilon)] = Var[X^\intercal \theta^* + X^\intercal (XX^\intercal)^{-1}X\varepsilon] = \ ...$$
$$... = Var[X^\intercal (XX^\intercal)^{-1}X\varepsilon] = X^\intercal (XX^\intercal)^{-1}XVar(\varepsilon)X^\intercal (XX^\intercal)^{-1}X = \Pi D \Pi^\intercal$$

where $\Pi = X^\intercal (XX^\intercal)^{-1}X = \Pi^\intercal$, and $D = diag\{\sigma_1^2, \sigma_2^2, ..., \sigma_n^2\}$.

Now suppose additionally that $\sigma_i^2 \equiv \sigma^2 > 0$. What can be said about distribution of the estimator $\hat{\sigma^2}$?

Insert solution here

Exercise 3. Consider the linear regression model from exercise 1. Suppose, that the target of estimation is $h^\intercal\theta$ for some determinate non-zero vector $h \in R^p$. Find expression for the LSE of $h^\intercal\theta$. Is this estimate optimal in sense of Gauss-Markov theorem, i.e. does it have the smallest variance among all linear unbiased estimators?

—Start with this —By Gauss Markov, we know that a BLUE estimator has $Var(\theta_{OLS}) = \sigma^2(XX^\intercal)^{-1})$. However in the case of heterscedastic noise, we have $Var(\theta) = (XX^\intercal)^{-1}XDX^\intercal(XX^\intercal)^{-1}$, which must be greater than $\sigma^2(XX^\intercal)^-1)$. An so, in this case, our estimator is not BLUE. Study the same issue for the target $\eta = H^\intercal\theta$, 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: $||Ax|| = ||x|| \;\forall\; x \in R^n \iff (Ax)^\intercal(Ay) \;\forall\; x, y \in R^n$.

$$||x|| = \sqrt{x \cdot x} = \sqrt{x^\intercal x} \implies ||Ax|| = \sqrt{Ax \cdot Ax} = \sqrt{x^\intercal A^\intercal Ax} \implies$$
$$A^\intercal A = I_n = A^{-1},\; A^\intercal = A^{-1}, ||Ax|| = ||x||$$

this implies A is an orthogonal matrix, and further,

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

Exercise 5. (a) Let $x_0 \in R^n$ be some fixed vector, find a projection map on the subspace $span(x_0)$. 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) $|PL2| \leq |PL1| \;\forall x \in R^n$, (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 $X \sim N(0, In)$, $Q = XX$. Suppose that Q is decomposed into the sum of two quadratic forms: $Q = Q1 + Q2$, where $Qi = X^\intercal A_i X$, $i = 1, 2$ for some symmetric matrices $A1, A2$ with $rank(A1) = n1$ and $rank(A2) = n2$. Show that if $n1 + n2 = n$, then $Q1$ and $Q2$ 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 $\eta = H^\intercal\theta^*$, where $H \in R^{q\times p}$ is some non-zero matrix with $q \leq p$. Find an analogue of the quadratic form $S2$ (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 $p = 2, X_i = (1, x_i)^\intercal, \theta^* = (\theta_1^*, \theta_2^*)^\intercal$ (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 $E[Y_1], ..., E[Y_n]$ in model (3).