

Homework 4

Due March 8 at 11 pm

1. (Condition number) Let $A \in \mathbb{R}^{n \times n}$ be invertible, and let $x_{\text{true}}, y \in \mathbb{R}^n$ satisfy $Ax_{\text{true}} = y$. We are interested in what happens if y is perturbed additively by a vector $z \in \mathbb{R}^n$, i.e. if we solve

$$Aw = y + z. \quad (1)$$

- (a) The operator norm of a matrix M is equal to

$$\|M\| := \max_{\|v\|_2=1} \|Mv\|_2, \quad (2)$$

which we know is equal to the maximum singular value. What is the operator norm of A^{-1} ?

- (b) Prove that $\|w - x_{\text{true}}\|_2 \leq \|z\|_2 / s_n$, where s_j denotes the j th singular value of A .

- (c) If $x_{\text{true}} \neq 0$ prove that

$$\frac{\|w - x_{\text{true}}\|_2}{\|x_{\text{true}}\|_2} \leq \kappa(A) \frac{\|z\|_2}{\|y\|_2}.$$

Here $\kappa(A) := s_1/s_n$ is called the *condition number* of A .

2. (Simple linear regression) We consider a linear model with one feature ($p := 1$). The data are given by

$$\tilde{y}_i := x_i \beta + \tilde{z}_i, \quad 1 \leq i \leq n, \quad (3)$$

where $\beta \in \mathbb{R}$, $x_i \in \mathbb{R}$, and $\tilde{z}_1, \dots, \tilde{z}_n$ are iid Gaussian random variables with zero mean and variance σ^2 . A reasonable definition of the *energy* in the feature is its sample mean square $\gamma^2 := \frac{1}{n} \sum_{i=1}^n x_i^2$. We define the signal-to-noise ratio in the data as $\text{SNR} := \gamma^2 / \sigma^2$.

- (a) What is the distribution of the OLS estimate $\tilde{\beta}_{OLS}$ as a function of the SNR?
- (b) If the SNR is fixed, how does the estimate behave as $n \rightarrow \infty$? If n is fixed, how does the estimate behave as $\text{SNR} \rightarrow \infty$? Can this behavior change if the noise is iid, has zero mean and variance σ^2 , but is not Gaussian? Prove that it doesn't or provide an example where it does.
- (c) Can the behavior of the estimator as $n \rightarrow \infty$ change if the noise is not iid? Prove that it doesn't or provide a counterexample.

3. (Best unbiased estimator) Consider the linear regression model

$$\tilde{y} = X^T \beta + \tilde{z}$$

where $\tilde{y} \in \mathbb{R}^n$, $X \in \mathbb{R}^{p \times n}$ has rank p , $\beta \in \mathbb{R}^p$, and $\tilde{z} \in \mathbb{R}^n$ has mean zero and covariance matrix $\Sigma_z = \sigma^2 I$ for some $\sigma^2 > 0$. Here only \tilde{z} and \tilde{y} are random. We observe the values of \tilde{y} and X and must estimate β . Consider a linear estimator of the form $C\tilde{y}$ where $C \in \mathbb{R}^{p \times n}$ (note that X and C are both deterministic, i.e., not random).

- (a) What is the mean $\mu = E[C\tilde{y}]$?
- (b) What is the covariance matrix of $C\tilde{y}$? That is, compute

$$E[(C\tilde{y})(C\tilde{y})^T] - \mu\mu^T.$$

- (c) Write $C = (XX^T)^{-1}X + D$ for some $D \in \mathbb{R}^{p \times n}$. What must be true of D so that $C\tilde{y}$ is an unbiased estimator of β for all possible β ? That is, what must be true so that $E[C\tilde{y}] = \beta$ for all β ? [Hint: Use part (a). Your answer will be a property of DX^T .]
- (d) Let Σ_C denote the covariance matrix of $C\tilde{y}$ and let Σ_{OLS} denote the covariance matrix of $(XX^T)^{-1}X\tilde{y}$. Show that if $C\tilde{y}$ is an unbiased estimator of β then

$$v^T \Sigma_C v \geq v^T \Sigma_{OLS} v,$$

for all $v \in \mathbb{R}^p$. That is, least squares yields the estimator with smallest variance in any direction v . [Hint: Use part (b) to compute the covariance of $((XX^T)^{-1}X + D)\tilde{y}$.]

- (e) Now suppose that the true regression model has extra features:

$$\tilde{y} = X^T\beta + Z^Tw + \tilde{z},$$

where $Z \in \mathbb{R}^{k \times n}$ and $w \in \mathbb{R}^k$. Not knowing these features, you compute the least squares estimator

$$\hat{\beta} = (XX^T)^{-1}X\tilde{y}.$$

Under what conditions on X, Z is $\hat{\beta}$ still unbiased for all possible w ?

4. (Distribution of β) In this question, we will investigate how the coefficients of regression, β is distributed. We will use the [combined cycle power plant data set](#) to regress for the net hourly electrical energy output as a function of the ambient temperature and exhaust vacuum. The support code loads the datasets and defines these subset of variables as X and y respectively. We will fit a regression to obtain β_0, β which minimizes $y = \beta_0 + \beta^T x$. To study the distribution of β , we split our dataset into 500 bootstrap samples, each with 100 data points. We fit linear regression individually on each of these 500 bootstrap samples to obtain $\beta^1, \beta^2, \dots, \beta^{500}$.

- (a) Plot a histogram of the distribution of β_1^k and β_2^k where k refers to the k^{th} bootstrap sample and β_i refers to the i^{th} component of β . The support code handles the actual plotting part, you only have to compute the β^k s.
- (b) Make a scatter plot of β_1^k vs β_2^k . Plot the principal directions of the actual data X and the principal directions of β^k s.
- (c) Do the principal directions of X datapoints and β^k datapoints align? Give a condition on the data generation process under which these principal directions will align.