MATH 156: Homework 1

Due on Tuesday, April 16

1. Consider the data: $Y_i = \beta_0 + \beta_1 + e_i$ where $e_1, \ldots, e_n$ are uncorrelated errors with mean zero and variance 1.

(a) Write this model in the form $Y = X\beta + e$ with $\beta = (\beta_1, \beta_2)$. Is the parameter β_1 estimable ? Why ?

(b) What is the least squares estimate of $\beta_0 + \beta_1$?

(c) Consider now a new observation $Y_{n+1} = \beta_0 + 2\beta_1 + e_{n+1}$ where $e_1, \ldots, e_{n+1}$ are uncorrelated errors with mean zero and variance 1. Write this model in the form $Y = X\beta + e$ and calculate the least squares estimate of β.

(d) Assume that $e_1 \ldots e_{n+1}$ are i.i.d. $\mathcal{N}(0,1)$. What is the MLE of β ?

(e) Take $\beta_0 = 2$, $\beta_1 = 3$, and $e_1 \ldots e_{n+1}$ i.i.d. $\mathcal{N}(0,1)$. Simulate $Y = (Y_1, \ldots, Y_{n+1})$, and plot the least squares values of β_0, β_1 for $n = 50, 100, 150, 200, \ldots, 1000$. What do you conclude ?

2. Suppose there are 4 objects whose individual weights $\beta_1, \ldots, \beta_4$ need to be estimated. We have a weighing balance which gives measurements with error having mean zero and variance σ^2. One approach is to weigh each object a number of times and take the average measurement value as the estimate of its weight. Such a procedure needs a total of 32 weighings (8 for each of the 4 objects) to estimate the weight of each object with precision (variance) $\sigma^2/8$.

Another method is to weigh the objects in combinations. Each operation consists in placing some of the objects in one pan of the balance and the others in the other pan. One then places some weights in the two pans to achieve equilibrium. This results in an observational equation of the type

$$y = x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + e,$$

where x_i is $0, 1$ or -1 according as the i^{th} object is not used, placed in the left pan or in the right pan of the balance and y is the weight required for equilibrium. After n measurements, one can get data that can be represented in an $n \times 1$ vector Y and an $n \times 4$ matrix X.

(a) Assume $n = 8$, and

$$X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

What are the least squares estimates of β_1, β_2, β_3 and β_4 ?

(b) For X as above, what is the covariance matrix of $\widehat{\beta}$? How many weightings does this scheme take to estimate all the weights with individual precision $\sigma^2/8$? Conclude.

(c) Does there exist a scheme of designing the 8×4 matrix X (each of whose elements is one among $0, 1$ and -1) so that the variance of any of the 4 weight estimates is strictly smaller than $\sigma^2/8$? Why or why not?

3. This problem focuses on the collinearity problem.

(a) Perform the following commands in R:

```
> set.seed(240)
> x1=runif(100)
> x2=0.5*x1+rnorm(100)/10
> y=2+2*x1+0.3*x2+rnorm(100)
```

The last line corresponds to creating a linear model in which y is a function of x_1 and x_2. Write out the form of the linear model. What are the regression coefficients?

(b) What is the correlation between x_1 and x_2? Create a scatterplot displaying the relationship between the variables.

(c) Using this data, fit a least squares regression to predict y using x_1 and x_2. Describe the results obtained. What are $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$? How do these relate to the true β_0, β_1 and β_2? Can you reject the null hypothesis $H_0 : \beta_1 = 0$? How about the null hypothesis $H_0 : \beta_2 = 0$?

(d) Now fit a least squares regression to predict y using only x_1. Comment on your results. Can you reject the null hypothesis $H_0 : \beta_1 = 0$?

(e) Now fit a least squares regression to predict y using only x_2. Comment on your results. Can you reject the null hypothesis $H_0 : \beta_1 = 0$?

(f) Do the results obtained in (c)-(e) contradict each other? Explain you answer.

(g) Now suppose we obtain one additional observation, which was unfortunately mismeasured.

```
> x1=c(x1, 0.1)
> x2=c(x2, 0.8)
> y=c(y,6)
```

Re-fit the linear models from (c) to (e) using this new data. What effect does this new observation have on the each of these models? In each model, is this observation an outlier? A high-leverage point? Both? Explain your answers.

4. Consider the linear regression model for which $\mathbb{E}[Y|X] = X\beta$ and $Cov[Y|X] = \sigma^2\mathbf{I}_n$, where $Y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times p}$. Please derive the closed-form solutions for the following optimization problems.

(a) Ordinary Least Squares (OLS).

$$\min_{\beta} ||Y - X\beta||_2^2.$$

(b) Ridge.

$$\min_{\beta} ||Y - X\beta||_2^2 + \lambda||\beta||_2^2.$$

(c) LASSO under the orthonormal covariates, i.e., $X^T X = \mathbf{I}_n$.

$$\min_{\beta} ||Y - X\beta||_2^2 + \lambda||\beta||_1.$$

5. We are going to explore the property of Elastic net by a simulation study.

(a) Consider the data structure $(X, Y) \sim P$, where $Y \in \mathbb{R}$ is a scalar outcome and $X = (X_j : j = 1, \dots, J) \in \mathbb{R}^J$, a J-dimensional vector of covariates. Assume the following Gaussian linear regression model

$$Y|X \sim \mathcal{N}(X^T\beta, \sigma^2) \text{ and } X \sim \mathcal{N}(0_{J\times 1}, \Gamma)$$

where $0_{J\times 1}$ is a J-dimensional column vector of zeros and the covariance matrix $\Gamma = (\gamma_{j,j'} : j, j' = 1, \dots, J)$ of the covariates has an autocorrelation of order 1, i.e., AR(1), structure,

$$\gamma_{j,j'} = \rho^{|j-j'|}$$

for $\rho \in (-1, 1)$. Set the parameter values to $J = 10, \beta = (-J/2+1, \dots, -2, -1, 0, 0, 1, 2, \dots, J/2-1)/10$, $\sigma = 2$, and $\rho = 0.5$. Simulate a learning set $\mathcal{L}_n = \{(X_i, Y_i) : i = 1, \dots, n\}$ of $n = 100$ i.i.d. random variables $(X_i, Y_i) \sim P, i = 1, \dots, n$. Also simulate an independent test set $\mathcal{T}_{n_{TS}} = \{(X_i, Y_i) : i = 1, \dots, n_{TS}\}$ of $n_{TS} = 1000$ i.i.d $(X_i, Y_i) \sim P, i = 1, \dots, n_{TS}$. Provide numerical and graphical summaries of the simulation model and of the learning set.

(b) The elastic net estimator of the regression coefficients β is defined as

$$\begin{aligned}
\hat{\beta}_n^{enet} &\equiv \text{argmin}_{\beta \in \mathbb{R}^J} ||Y - X\beta||_2^2 + \lambda_1 ||\beta||_1 + \lambda_2 ||\beta||_2^2 \\
&= \text{argmin}_{\beta \in \mathbb{R}^J} \sum_{i=1}^{n} \left(Y_i - \sum_{j=1}^{J} \beta_j X_{i,j} \right)^2 \\
&\quad + \lambda_1 \sum_{j=1}^{J} |\beta_j| + \lambda_2 \sum_{j=1}^{J} \beta_j^2,
\end{aligned}$$

where the shrinkage parameters $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ are tuning parameters that control the strength of the penalty terms, i.e., the complexity or shrinking of the coefficients towards zero. Obtain ridge ($\lambda_1 = 0, \lambda_2 = \lambda$), LASSO ($\lambda_1 = \lambda, \lambda_2 = 0$), and elastic net ($\lambda_1 = \lambda_2 = \lambda/2$) estimators of the regression coefficients β, for $\lambda \in \{0, 1, ..., 100\}$, based on the learning set simulated in (a). In particular, for each type of estimator, provide and comment on plots of the effective degrees of freedom versus the shrinkage parameter λ and plots of the estimated regression coefficients versus the shrinkage parameter. For each type of estimator, obtain the learning set risk for the squared error loss function, i.e., the mean squared error (MSE),

$$MSE(\hat{\beta}; \mathcal{L}_n) = \frac{1}{n} ||Y - X\hat{\beta}||_2^2.$$

Provide and comment on plots of the MSE versus the shrinkage parameter and report which values of the shrinkage parameter minimize risk.
Hint. You may use the *glmnet* function from the glmnet package, but be mindful of centering and scaling, of the handling of the intercept, and of the parameterization of the elastic net penalty.

(c) For each estimator in (b), obtain the test set risk $MSE(\hat{\beta}; \mathcal{T}_{n_{TS}})$ for the squared error loss function (i.e., MSE). Provide and comment on plots of risk versus the shrinkage parameter and report which values of the shrinkage parameter minimize risk. Examine the corresponding three "optimal" estimators of the regression coefficients.