STAT 7650. Homework 1 (Due: Tuesday, January 27, 2026)

1. (35 points). Each female horseshoe crab has a male crab residing in her nest. Satellite males are additional male crabs that reside nearby. Let the response variable be $Y = 1$ if the female crab has at one satellite male and $Y = 0$ otherwise. The predictor variable is the width of the female crab. In the data file, `satell` denotes the number of satellite males and `width` denote the female crab's width.

   `http://www.auburn.edu/~zengpen/teaching/STAT-7030/datasets/crab.txt`

   Consider a probit regression model,

   $$\pi_i = P(Y_i = 1 \mid X = x_i) = \Phi(\beta^T x_i), \quad i = 1, \ldots, n,$$

   where $\Phi(\cdot)$ denotes the CDF of $N(0, 1)$ and $x = (1, \text{width})^T$. More generally, denote the observed data as $\{(y_i, x_i), y_i \in \{0, 1\}, x_i \in \mathbb{R}^p, i = 1, \ldots, n\}$.

   (a) Write down the log-likelihood function for the probit model. Implement a function to compute the log-likelihood, with input arguments consisting of the coefficient vector $\beta \in \mathbb{R}^p$, the response vector $y \in \{0, 1\}^n$, and the design matrix $X \in \mathbb{R}^{n \times p}$. Evaluate the log-likelihood for the crab data at $\beta = (-7.5, 0.3)^T$.

   ```
   probit_loglik = function(b, y, X) { ... }
   ```

   (b) Derive the gradient of the log-likelihood with respect to β.

   (c) Implement a function to compute the gradient of the log-likelihood. Evaluate the gradient for the crab data at $\beta = (-7.5, 0.3)^T$.

   ```
   probit_grad = function(b, y, X) { ... }
   ```

   (d) Implement a function to compute the MLE of β using `optim()` with the BFGS method. Make sure to specify functions to compute the log-likelihood and its gradient properly. You may set the initial value as 0.

   ```
   fit_probit_optim = function(b, y, X) { ... }
   ```

   (e) Derive the observed information matrix $J_n(\theta)$.

   (f) Implement a function to compute the observed information matrix. Evalute $J_n(\theta)$ for the crab data at the MLE of β.

   ```
   probit_hessian = function(b, y, X) { ... }
   ```

   (g) Find the standard error of the MLE of β for the crab data based on the observed information matrix.

2. (20 points) The observations $\{(y_i, x_i), y_i \in \mathbb{R}, x_i \in \mathbb{R}^p, i = 1, \dots, n\}$ are assumed to be independently drawn from a linear regression model

$$y_i = \beta^T x_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \quad i = 1, \dots, n,$$

where $\beta \in \mathbb{R}^p$ is a vector of regression coefficients. The parameter vector to be estimated is $\theta = (\beta^T, \sigma^2)^T$.

(a) Write down the log-likelihood function for the linear regression model and derive its gradient with respect to the parameter vector θ.

(b) Obtain the MLE of β and σ^2. Discuss the most efficient method for computing the MLEs.

(c) Derive the Fisher information matrix

$$I(\theta) = E\left[-\frac{\partial^2 \log f(\theta)}{\partial\theta\partial\theta^T}\right],$$

where $f(\theta)$ indicates the density of a single observation.

(d) Implement a function to compute the MLE of β and σ^2, along with their standard errors. Compute the standard errors based on the Fisher information matrix evaluated at the MLE.

```
fit_lm_mle = function(y, x) {
    ...
    list(b = , sigma2 = , b_sd = , sigma2_sd = )
}
```

3. (25 points) The observations $\{(y_i, x_i), y_i \in \mathbb{R}, x_i \in \mathbb{R}^p, i = 1, \dots, n\}$ are assumed to be independently drawn from a linear regression model

$$y_i = \beta^T x_i + \varepsilon_i, \quad i = 1, \dots, n,$$

where $\beta \in \mathbb{R}^p$ is a vector of regression coefficients. For a given constant $\lambda > 0$, ridge regression estimates β by solving

$$\hat{\beta}_\lambda = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} (y_i - \beta^T x_i)^2 + \lambda \beta^T \beta \right\}.$$

(a) Derive an explicit expression for the ridge regression estimator $\hat{\beta}_\lambda$.

(b) Discuss how $\hat{\beta}_\lambda$ can be computed efficiently using the SVD-decomposition of X.

(c) Implement a function to compute the ridge regression estimator $\hat{\beta}_\lambda$.

```
fit_lm_ridge = function(y, x, lambda) { ... }
```

(d) (10 points) Conduct a small simulation study to investigate the effect of λ on the performance of ridge regression.

- Generate a dataset as follows.
  - Set $n = 100$, $p = 5$, and $\beta_* = (2, 0, -1, 0.5, 0)^T$.
  - Generate the components of x_i and ε_i independently from $N(0, 1)$.
  - Compute $y_i = \beta^T x_i + \varepsilon_i$, for $i = 1, \ldots, n$.
- Fit a ridge regression model using a specified value of λ.
- Evaluate the performance of the estimators by $D(\hat{\beta}_\lambda, \beta_*) = \|\hat{\beta} - \beta_*\|_2^2$.
- Repeat the above steps 200 times and summarize $D(\hat{\beta}_\lambda, \beta_*)$ using mean and the 95% confidence interval for mean.
- Repeat the entire procedure for different values of $\lambda = 0, e^{-5}, e^{-4}, e^{-3}, \ldots, e^2$ and plot the resulting means with error bars as a function of λ, where the error bars indicate 95% confidence intervals.

Discuss the effect of λ on the performance of ridge regression based on the results of your simulation study.

4. (20 points) Implement functions for the following problems. Try your best to make your codes efficient.

(a) For a symmetric matrix A, compute a symmetric matrix $A^{1/2}$ such that $A = A^{1/2}A^{1/2}$.

```
mat_sqrt = function(A) { ... }
```

(b) For a positive definite matrix A, compute a symmetric matrix $A^{-1/2}$ such that $A^{-1/2}AA^{-1/2} = I$.

```
mat_sqrt_inv = function(A) { ... }
```

(c) For a matrix X and a vector b, compute $X(X^TX)^{-1}X^Tb$, which is the projection of b onto the column space of X.

```
proj_span = function(x, b) { ... }
```

(d) For a matrix $A \in \mathbb{R}^{m \times n}$ of full column rank and $m > n$, compute its polar decomposition $H = A(A^TA)^{-1/2}$ and $Z = (A^TA)^{1/2}$.

```
polar = function(A) { ...; list(orient = , scale = ); }
```