

# Breast Cancer Diagnosis

Hongjie Liu, Xicheng Xie, Jiajun Tao, Shaohan Chen, Yujia Li

April 3rd, 2023

# Outline

- ▶ Background
- ▶ Task Introduction
- ▶ Task 1
- ▶ Task 2
- ▶ Task 3
- ▶ Task 4
- ▶ Discussions
- ▶ Reference
- ▶ Q&A

# Background

## **Breast Cancer Diagnosis:**

- ▶ In this project we study the breast cancer diagnosis problem
- ▶ The goal of the exercise is to build a predictive model based on logistic regression to facilitate cancer diagnosis
- ▶ We move towards the goal with the steps of task 1, 2, 3 and 4.

# Background

## Data Source:

- ▶ The data is the breast cancer medical data retrieved from “breast-cancer.csv”, which has 569 rows and 32 columns
- ▶ The first column ID labels individual breast tissue images; The second column Diagnosis identifies if the image is coming from cancer tissue or benign cases (M = malignant, B = benign). There are 357 benign and 212 malignant cases
- ▶ The other 30 columns correspond to mean, standard deviation and the largest values (points on the tails) of the distributions of the following 10 features computed for the cell nuclei

# Task Introduction

- ▶ Task 1: Build a logistic model to classify the images into malignant/benign, and write down your likelihood function, its gradient and Hessian matrix.
- ▶ Task 2: Develop a Newton-Raphson algorithm to estimate your model.
- ▶ Task 3: Build a logistic-LASSO model to select features, and implement a path-wise coordinate-wise optimization algorithm to obtain a path of solutions with a sequence of descending  $\lambda$ 's.
- ▶ Task 4: Use 5-fold cross-validation to select the best  $\lambda$ . Compare the prediction performance between the 'optimal' model and 'full' model

# Task 1 - Objective

## **Objective:**

Build a logistic model to classify the images into malignant/benign, and write down your likelihood function, its gradient and Hessian matrix.

# Task 1 - Build a logistic model

Define the “Diagnosis” variable will be coded as 1 for malignant cases and 0 for benign cases.

Given  $n$  i.i.d. observations with  $p$  predictors, we consider a logistic regression model

$$P(Y_i = 1 \mid \mathbf{x}_i) = \frac{e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}, \quad i = 1, \dots, n \quad (1)$$

where  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top \in \mathbb{R}^{p+1}$  is the parameter vector,  $\mathbf{x}_i = (1, X_{i1}, \dots, X_{ip})^\top$  is the vector of predictors in the  $i$ -th observation, and  $Y_i \in \{0, 1\}$  is the binary response in the  $i$ -th observation.

## Task 1 - Build a logistic model

Let  $\mathbf{y} = (Y_1, Y_2, \dots, Y_n)^\top$  denote the response vector, and  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^\top \in \mathbb{R}^{n \times (p+1)}$  denote the design matrix.

The observed likelihood of  $\{(Y_1, \mathbf{x}_1), (Y_2, \mathbf{x}_2), \dots, (Y_n, \mathbf{x}_n)\}$  is

$$L(\beta; \mathbf{y}, \mathbf{X}) = \prod_{i=1}^n \left[ \left( \frac{e^{\mathbf{x}_i^\top \beta}}{1 + e^{\mathbf{x}_i^\top \beta}} \right)^{Y_i} \left( \frac{1}{1 + e^{\mathbf{x}_i^\top \beta}} \right)^{1-Y_i} \right]$$



## Task 1 - Build a logistic model

Maximizing the likelihood is equivalent to maximizing the log-likelihood function:

$$f(\beta; \mathbf{y}, \mathbf{X}) = \sum_{i=1}^n \left[ Y_i \mathbf{x}_i^\top \beta - \log \left( 1 + e^{\mathbf{x}_i^\top \beta} \right) \right]. \quad (2)$$

The estimates of model parameters are

$$\hat{\beta} = \arg \max_{\beta} f(\beta; \mathbf{y}, \mathbf{X}),$$

and the optimization problem is

$$\max_{\beta} f(\beta; \mathbf{y}, \mathbf{X}). \quad (3)$$

## Task 1 - Build a logistic model

Denote  $p_i = P(Y_i = 1 \mid \mathbf{x}_i)$  as given in (1) and  $\mathbf{p} = (p_1, p_2, \dots, p_n)^\top$ . The gradient of  $f$  is:

$$\begin{aligned}\nabla f(\boldsymbol{\beta}; \mathbf{y}, \mathbf{X}) &= \mathbf{X}^\top (\mathbf{y} - \mathbf{p}) \\ &= \sum_{i=1}^n (Y_i - p_i) \mathbf{x}_i \\ &= \begin{pmatrix} \sum_{i=1}^n (Y_i - p_i) \\ \sum_{i=1}^n (Y_i - p_i) X_{i1} \\ \vdots \\ \sum_{i=1}^n (Y_i - p_i) X_{ip} \end{pmatrix}\end{aligned}$$

## Task 1 - Build a logistic model

Denote  $w_i = p_i(1 - p_i) \in (0, 1)$  and  $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$ . The Hessian matrix of  $f$  is given by

$$\begin{aligned}\nabla^2 f(\beta; \mathbf{y}, \mathbf{X}) &= -\mathbf{X}^\top \mathbf{W} \mathbf{X} \\ &= -\sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\top \\ &= -\begin{pmatrix} \sum_{i=1}^n w_i & \sum_{i=1}^n w_i X_{i1} & \cdots & \sum_{i=1}^n w_i X_{i1} \\ \sum_{i=1}^n w_i X_{i1} & \sum_{i=1}^n w_i X_{i1}^2 & \cdots & \sum_{i=1}^n w_i X_{i1} X_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n w_i X_{ip} & \sum_{i=1}^n w_i X_{ip} X_{i1} & \cdots & \sum_{i=1}^n w_i X_{ip}^2 \end{pmatrix}\end{aligned}$$

## Task 1 - Build a logistic model

Next, we show that the Hessian matrix  $\nabla^2 f(\beta; \mathbf{y}, \mathbf{X})$  is a negative-definite matrix if  $\mathbf{X}$  has full rank.

**Proof.** For any  $(p + 1)$ -dimensional nonzero vector  $\alpha$ , given that  $\mathbf{X}$  has full rank,  $\mathbf{X}\alpha$  is also a nonzero vector. Since  $\mathbf{W}$  is positive-definite, we have

$$\begin{aligned}\alpha^\top \nabla^2 f(\beta; \mathbf{y}, \mathbf{X}) \alpha &= \alpha^\top (-\mathbf{X}^\top \mathbf{W} \mathbf{X}) \alpha \\ &= -(\mathbf{X}\alpha)^\top \mathbf{W} (\mathbf{X}\alpha) \\ &< 0.\end{aligned}$$

Thus,  $\nabla^2 f(\beta; \mathbf{y}, \mathbf{X})$  is negative-definite. □

Hence, the optimization problem (3) is a well-defined problem.

## Task 2 - Objective

### Objective:

Develop a Newton-Raphson algorithm to estimate your model

- ▶ Recall The target function  $f$  given in task 1 is:

$$f(\beta; \mathbf{y}, \mathbf{X}) = \sum_{i=1}^n \left[ Y_i \mathbf{x}_i^\top \beta - \log \left( 1 + e^{\mathbf{x}_i^\top \beta} \right) \right].$$

- ▶ We develop a modified Newton-Raphson algorithm including a step-halving step to maximize the target function.

## Task 2 - Newton-Raphson algorithm design

---

### Algorithm 1 Newton-Raphson algorithm

---

**Require:**  $f(\beta)$  - target function as given in (2);  $\beta_0$  - starting value

**Ensure:**  $\hat{\beta}$  such that  $\hat{\beta} \approx \arg \max_{\beta} f(\beta)$

$i \leftarrow 0$ , where  $i$  is the current number of iterations

$f(\beta_{-1}) \leftarrow -\infty$

**while** convergence criterion is not met **do**

$i \leftarrow i + 1$

$\mathbf{d}_i \leftarrow -[\nabla^2 f(\beta_{i-1})]^{-1} \nabla f(\beta_{i-1})$ , where  $\mathbf{d}_i$  is the direction in the  $i$ -th iteration

$\lambda_i \leftarrow 1$ , where  $\lambda_i$  is the multiplier in the  $i$ -th iteration

$\beta_i \leftarrow \beta_{i-1} + \lambda_i \mathbf{d}_i$

**while**  $f(\beta_i) \leq f(\beta_{i-1})$  **do**

$\lambda_i \leftarrow \lambda_i / 2$

$\beta_i \leftarrow \beta_{i-1} + \lambda_i \mathbf{d}_i$

**end while**

**end while**

$\hat{\beta} \leftarrow \beta_i$

---

## Task 2 - Newton-Raphson algorithm R code

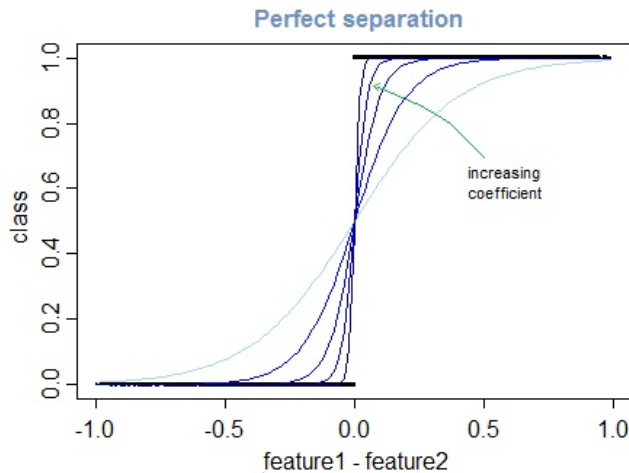
```
NewtonRaphson <- function(dat, func, start, tol = 1e-8, maxiter = 200) {  
  i <- 0  
  cur <- start  
  stuff <- func(dat, cur)  
  res <- c(0, stuff$f, cur)  
  prevf <- -Inf  
  X <- cbind(rep(1, nrow(dat)), as.matrix(dat[, -1]))  
  y <- dat[, 1]  
  warned <- 0  
  while (abs(stuff$f - prevf) > tol && i < maxiter) {  
    i <- i + 1  
    prevf <- stuff$f  
    prev <- cur  
    d <- -solve(stuff$Hess) %*% stuff$grad  
    cur <- prev + d  
    lambda <- 1  
    maxhalv <- 0  
    while (func(dat, cur)$f < prevf && maxhalv < 50) {  
      maxhalv <- maxhalv + 1  
      lambda <- lambda / 2  
      cur <- prev + lambda * d  
    }  
    stuff <- func(dat, cur)  
    res <- rbind(res, c(i, stuff$f, cur))  
    y_hat <- ifelse(X %*% cur > 0, 1, 0)  
    if (warned == 0 && sum(y - y_hat) == 0) {  
      warning("Complete separation occurs. Algorithm does not converge.")  
      warned <- 1  
    }  
  }  
  colnames(res) <- c("iter", "target_function", "(Intercept)", names(dat)[-1])  
  return(res)  
}
```

## Task 2 - Complete separation

- ▶ Sometimes our algorithm does not converge because of the complete separation.
- ▶ A complete separation in a logistic regression, sometimes also referred as perfect prediction, occurs whenever there exists some vector of coefficients  $\beta$  such that  $Y_i = 1$  whenever  $\mathbf{x}_i^\top \beta > 0$  and  $Y_i = 0$  whenever  $\mathbf{x}_i^\top \beta \leq 0$ .
- ▶ Complete separation occur when a linear function of predictors can perfectly classify the response.



## Task 2 - Complete separation



## Task 2 - Complete separation

- ▶ We have proved that: when there exists a vector of coefficients  $\hat{\beta}$  such that  $Y_i = 1$  whenever  $\mathbf{x}_i^\top \hat{\beta} > 0$  and  $Y_i = 0$  whenever  $\mathbf{x}_i^\top \hat{\beta} \leq 0$ , there does not exist  $\beta^* \in \mathbb{R}^{(p+1)}$  such that  $\beta^* = \arg \max_{\beta} f(\beta)$ , where  $f$  is given in (2). Thus our algorithm does not converge. (proof is attached in report appendix)
- ▶ If there is no complete separation, the parameters output by glm function and our algorithm are demonstrated to be the same
- ▶ In this case, given the perfect separation does exist, we choose the  $\beta$  based on the pre-defined `tol = 1e-8` in the algorithm.
- ▶ The difference of  $\beta$  in our algorithm output and glm function is out of the different termination condition

## Task 2 - Comparison

Comparison of using glm function and our algorithm (part of)

predictor	ours	glm
(Intercept)	111.7230206	90.9690365
radius_mean	-3646.8235387	-2560.3938902
texture_mean	-2.8949199	0.8812037
perimeter_mean	1257.9481173	789.2724398
area_mean	2091.5001210	1539.8345650
smoothness_mean	180.4591375	128.8762247
compactness_mean	-471.3749334	-346.6691873
concavity_mean	-0.3063034	9.0810082
concave.points_mean	287.3555092	215.4808031
symmetry_mean	10.8459784	10.4772590
fractal_dimension_mean	-40.6413497	-28.6916549
radius_se	854.4290933	617.5687358
texture_se	-105.8920782	-79.9788638
perimeter_se	-1231.6514585	-917.3916527
area_se	789.0550146	628.6222313
smoothness_se	-83.3449730	-63.7660367
compactness_se	473.1970464	344.3180183
concavity_se	-440.6629325	-323.8533730
concave.points_se	503.3728692	374.7610875
symmetry_se	-144.4580558	-108.6211792

## Task 3 - Objective

### Objective:

Build a logistic-LASSO model to select features by implementing a path-wise coordinate-wise optimization algorithm

- Recall Log-likelihood  $f$  in task 1:

$$f(\beta; \mathbf{y}, \mathbf{X}) = \sum_{i=1}^n \left[ Y_i \mathbf{x}_i^\top \beta - \log \left( 1 + e^{\mathbf{x}_i^\top \beta} \right) \right].$$

- LASSO estimates the logistic model parameters  $\beta$  by optimizing a penalized loss function:

$$\min_{\beta} -\frac{1}{n} f(\beta) + \lambda \sum_{k=1}^p |\beta_k|. \quad (4)$$

where  $\lambda \geq 0$  is the tuning parameter. Note that the intercept is not penalized and all predictors are standardized.

## Task 3 - Algorithm of Logistic-LASSO Model

### Algorithm Structure:

- ▶ OUTER LOOP: Decrement  $\lambda$ .
- ▶ MIDDLE LOOP: Update  $\tilde{w}_i$ ,  $\tilde{p}_i$ , and thus the quadratic approximation  $\ell$  using the current parameters  $\tilde{\beta}$ .
- ▶ INNER LOOP: Run the coordinate descent algorithm on the penalized weighted-least-squares problem.

## Task 3 - OUTER LOOP

### OUTER LOOP:

Compute the solutions of the optimization problem (4) for a decreasing sequence of values for  $\lambda$ :  $\{\lambda_1, \dots, \lambda_m\}$ , starting at the smallest value  $\lambda_1 = \lambda_{max}$

$$\lambda_{max} = \frac{1}{n} \max_j |\langle \mathbf{x}_{.j}, \mathbf{y} \rangle|, \quad (5)$$

where  $\mathbf{x}_{.j}$  is the  $j$ -th column of the design matrix  $\mathbf{X}$ , for  $j = 1, \dots, p$ .

For tuning parameter value  $\lambda_{k+1}$ , we initialize coordinate descent algorithm at the computed solution for  $\lambda_k$  (warm start).

## Task 3 - MIDDLE LOOP

### MIDDLE LOOP:

- For a fixed  $\lambda$ , find the estimates of  $\beta$  by solving the optimization problem (4).

Based on current parameter estimates  $\tilde{\beta}$ , we form a quadratic approximation to the log-likelihood  $f$  using a Taylor expansion:

$$\begin{aligned} f(\beta) \approx \ell(\beta) &= f(\tilde{\beta}) + (\beta - \tilde{\beta})^\top \nabla f(\tilde{\beta}) + \frac{1}{2}(\beta - \tilde{\beta})^\top \nabla^2 f(\tilde{\beta})(\beta - \tilde{\beta}) \\ &= -\frac{1}{2} \sum_{i=1}^n \tilde{w}_i \left[ \mathbf{x}_i^\top (\tilde{\beta} - \beta) + \frac{Y_i - \tilde{p}_i}{\tilde{w}_i} \right] + \frac{1}{2} \sum_{i=1}^n \tilde{w}_i \left( \frac{Y_i - \tilde{p}_i}{\tilde{w}_i} \right)^2 + f(\tilde{\beta}), \end{aligned}$$

where  $\tilde{\mathbf{p}} = (\tilde{p}_1, \dots, \tilde{p}_n)^\top$  and  $\tilde{\mathbf{W}} = \text{diag}(\tilde{w}_1, \dots, \tilde{w}_n)$  are the estimates of  $\mathbf{p}$  and  $\mathbf{W}$  based on  $\tilde{\beta}$ .

## Task 3 - MIDDLE LOOP

- Let  $\tilde{z}_i = \mathbf{x}_i^\top \tilde{\beta} + \frac{Y_i - \tilde{p}_i}{\tilde{w}_i}$ , we have

$$\ell(\beta) = -\frac{1}{2} \sum_{i=1}^n \tilde{w}_i (\tilde{z}_i - \mathbf{x}_i^\top \beta)^2 + C(\tilde{\beta}), \quad (6)$$

where  $\tilde{z}_i$  is the working response,  $\tilde{w}_i$  is the working weight, and  $C$  is a function that does not depend on  $\beta$ .



## Task 3 - INNER LOOP

### INNER LOOP:

With fixed  $\tilde{w}_i$ 's,  $\tilde{z}_i$ 's, and a fixed form of  $\ell$  based on the estimates of  $\beta$  in the previous iteration of the middle loop, we use coordinate descent to update  $\beta$  by solving

$$\min_{\beta} -\frac{1}{n}\ell(\beta) + \lambda \sum_{k=1}^p |\beta_k|, \quad (7)$$

## Task 3 - INNER LOOP

- Coordinate-wise objective function Based the current estimates  $\tilde{\beta}_k$  for  $k \neq j$ :

$$\min_{\beta_j} \frac{1}{2n} \sum_{i=1}^n \tilde{w}_i \left( \tilde{z}_i - x_{ij}\beta_j - \sum_{k \neq j} x_{ik}\tilde{\beta}_k \right)^2 + \lambda |\beta_j| + \lambda \sum_{k \neq j} |\tilde{\beta}_k|.$$

- Updates:

$$\begin{aligned}\tilde{\beta}_0 &\leftarrow \frac{\sum_{i=1}^n \tilde{w}_i (\tilde{z}_i - \sum_{k=1}^p x_{ik}\tilde{\beta}_k)}{\sum_{i=1}^n \tilde{w}_i}, \\ \tilde{\beta}_j &\leftarrow \frac{S\left(\frac{1}{n} \sum_{i=1}^n \tilde{w}_i x_{ij} (\tilde{z}_i - \sum_{k \neq j} x_{ik}\tilde{\beta}_k), \lambda\right)}{\frac{1}{n} \sum_{i=1}^n \tilde{w}_i x_{ij}^2}, \quad j = 1, \dots, p\end{aligned}$$

where  $S(z, \gamma)$  is the soft-thresholding operator with value

$$S(z, \gamma) = \text{sign}(z)(|z| - \gamma)_+ = \begin{cases} z - \gamma, & \text{if } z > 0 \text{ and } \gamma < |z| \\ z + \gamma, & \text{if } z < 0 \text{ and } \gamma < |z| \\ 0, & \text{if } \gamma \geq |z| \end{cases}$$

Keep updating estimates of  $\beta_j$ 's repeatedly for  $j = 0, 1, 2, \dots, p, 0, 1, 2, \dots$  until convergence.

# Task 3 - Path-wise coordinate-wise optimization algorithm

---

**Algorithm 1** Path-wise coordinate-wise optimization algorithm

---

**Require:**  $g(\beta, \lambda) = -\frac{1}{n}f(\beta) + \lambda \sum_{k=1}^p |\beta_k|$  - target function, where  $f(\beta)$  is given in (1);  $\beta_0$  - starting value;  $\{\lambda_1, \dots, \lambda_m\}$  - a sequence of descending  $\lambda$ 's, where  $\lambda_1 = \lambda_{max}$  is given in (3);  $\epsilon$  - tolerance;  $N_s$ ,  $N_t$  - maximum number of iterations of the middle and inner loops

**Ensure:**  $\hat{\beta}(\lambda_r)$  such that  $\hat{\beta}(\lambda_r) \approx \arg \min_{\beta} g(\beta, \lambda_r)$ ,  $r = 1, \dots, m$

```
1:  $\tilde{\beta}_0(\lambda_1) \leftarrow \beta_0$ 
2: OUTER LOOP
3: for  $r \in \{1, \dots, m\}$ , where  $r$  is the current number of iterations of the outer loop, do
4:    $s \leftarrow 0$ , where  $s$  is the current number of iterations of the middle loop
5:    $g(\tilde{\beta}_{s-1}(\lambda_r), \lambda_r) \leftarrow \infty$ 
6:   MIDDLE LOOP
7:   while  $t \geq 2$  and  $s < N_s$  do
8:      $s \leftarrow s + 1$ 
9:     Update  $\tilde{w}_i^{(s)}$ ,  $\tilde{z}_i^{(s)}$  ( $i = 1, \dots, n$ ), and thus  $\ell_s(\beta)$  as given in (4) based on  $\tilde{\beta}_{s-1}(\lambda_r)$ 
10:     $t \leftarrow 0$ , where  $t$  is the current number of iterations of the inner loop
11:     $\tilde{\beta}_s^{(0)}(\lambda_r) \leftarrow \tilde{\beta}_{s-1}(\lambda_r)$ 
12:     $h_s(\tilde{\beta}_s^{(-1)}(\lambda_r), \lambda_r) \leftarrow \infty$ , where  $h_s(\beta, \lambda) = -\frac{1}{n}\ell_s(\beta) + \lambda \sum_{k=1}^p |\beta_k|$ 
13:    INNER LOOP
14:    while  $|h_s(\tilde{\beta}_s^{(t)}(\lambda_r), \lambda_r) - h_s(\tilde{\beta}_s^{(t-1)}(\lambda_r), \lambda_r)| > \epsilon$  and  $t < N_t$  do
15:       $t \leftarrow t + 1$ 
16:       $\tilde{\beta}_0^{(t)}(\lambda_r) \leftarrow \sum_{i=1}^n \tilde{w}_i^{(s)} \left( \tilde{z}_i^{(s)} - \sum_{k=1}^p x_{ik} \tilde{\beta}_k^{(t-1)}(\lambda_r) \right) / \sum_{i=1}^n \tilde{w}_i^{(s)}$ 
17:      for  $j \in \{1, \dots, p\}$  do
18:         $\tilde{\beta}_j^{(t)}(\lambda_r) \leftarrow S \left( \frac{1}{n} \sum_{i=1}^n \tilde{w}_i^{(s)} x_{ij} \left( \tilde{z}_i^{(s)} - \sum_{k < j} x_{ik} \tilde{\beta}_k^{(t)}(\lambda_r) - \sum_{k > j} x_{ik} \tilde{\beta}_k^{(t-1)}(\lambda_r) \right), \lambda_r \right) / \frac{1}{n} \sum_{i=1}^n \tilde{w}_i^{(s)} x_{ij}^2$ 
19:      end for
20:    end while
21:     $\tilde{\beta}_s(\lambda_r) \leftarrow \tilde{\beta}_s^{(t)}(\lambda_r)$ 
22:  end while
23:   $\hat{\beta}(\lambda_r) \leftarrow \tilde{\beta}_s(\lambda_r)$ 
24:   $\tilde{\beta}_0(\lambda_{r+1}) \leftarrow \hat{\beta}(\lambda_r)$ 
25: end for
```

---

## Task 4 - Objective

### Objective:

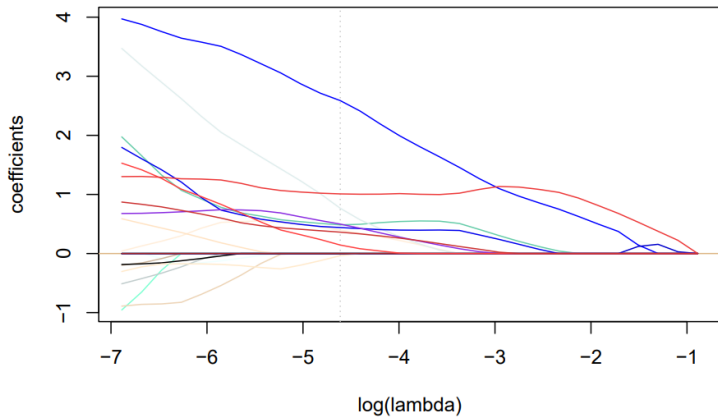
Use 5-fold cross-validation to select the best  $\lambda$ . Compare the prediction performance between the 'optimal' model and 'full' model.

- ▶ Select the best  $\lambda$ : our function `cv.logit.lasso` and `cv.glmnet`.
- ▶ Use the best lambda to select predictors and refit logistic regression model –'optimal' model.
- ▶ Compare the 'optimal' model and the 'full' model from Task 2 in Specificity, Sensitivity, and AUC.

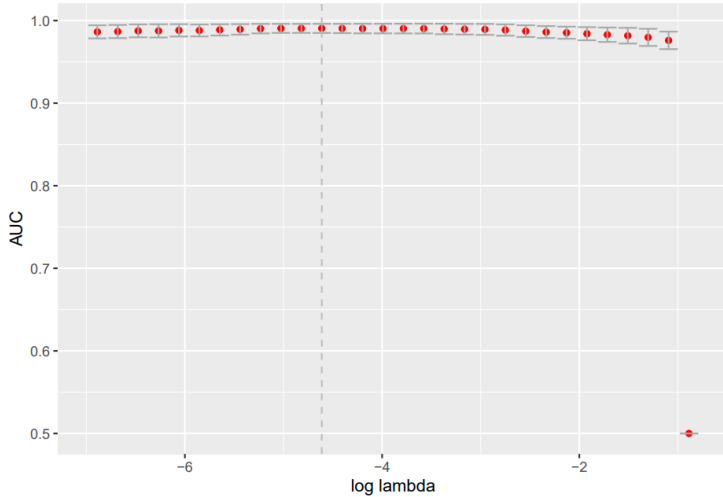
## Task 4 - Our CV function

```
cv.logit.lasso <- function(x, y, nfolds = 5, lambda) {  
  auc <- data.frame(matrix(ncol = 3, nrow = 0))  
  folds <- createFolds(y, k = nfolds)  
  for (i in 1:nfolds) {  
    valid_index <- folds[[i]]  
    x_training <- x[-valid_index, ]  
    y_training <- y[-valid_index]  
    training_dat <- data.frame(cbind(y_training, x_training))  
    x_valid <- cbind(rep(1, length(valid_index)), x[valid_index, ])  
    y_valid <- y[valid_index]  
    res <- LogisticLASSO(dat = training_dat, start = rep(0, ncol(training_dat)), lambda = lambda)  
    for (k in 1:nrow(res)) {  
      betavec <- res[k, 2:ncol(res)]  
      u_valid <- x_valid %*% betavec  
      phat_valid <- sigmoid(u_valid)[, 1]  
      roc <- roc(response = y_valid, predictor = phat_valid)  
      auc <- rbind(auc, c(lambda[k], i, roc$auc[1]))  
    }  
  }  
  colnames(auc) <- c("lambda", "fold", "auc")  
  cv_res <- auc %>%  
    group_by(lambda) %>%  
    summarize(auc_mean = mean(auc)) %>%  
    mutate(auc_ranking = min_rank(desc(auc_mean)))  
  bestlambda <- min(cv_res$lambda[cv_res$auc_ranking == 1])  
  return(cv_res)  
}
```

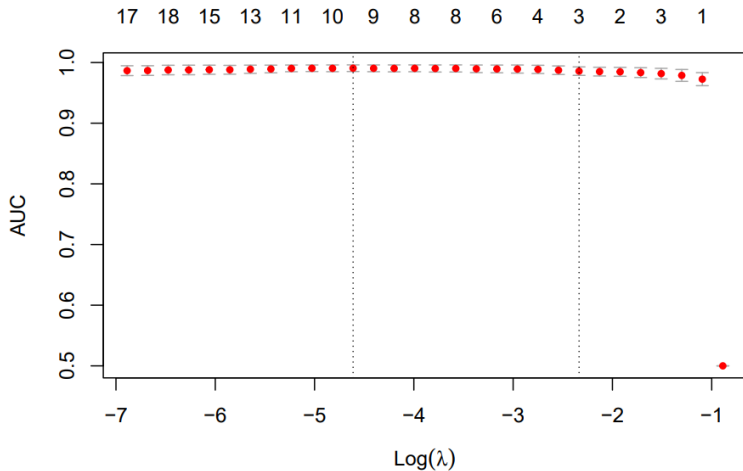
## Task 4 - Our CV function – coefficient shrinkage



## Task 4 - Our CV function – selecting the best lambda



## Task 4 - glmnet – selecting the best lambda





## Task 4 - Best lambda by the two methods

lambda	ours_AUC	cv.glmnet_AUC
0.4115445	0.5000000	0.5000000
0.3346284	0.9760054	0.9725915
0.2720876	0.9796678	0.9787526
0.2212355	0.9816903	0.9817053
0.1798874	0.9828852	0.9834002
0.1462671	0.9841071	0.9847216
0.1189303	0.9853029	0.9850192
0.0967027	0.9861031	0.9858166
0.0786293	0.9872060	0.9873215
0.0639338	0.9885875	0.9887026
0.0519848	0.9893408	0.9892575
0.0422691	0.9895397	0.9894566
0.0343691	0.9898394	0.9896556
0.0279457	0.9903360	0.9902522
0.0227227	0.9903354	0.9901527
0.0184759	0.9903346	0.9904508
0.0150229	0.9903347	0.9903518
0.0122151	0.9904332	0.9904501
0.0099322	0.9905325	0.9906491
0.0080759	0.9905317	0.9904501
0.0065665	0.9905324	0.9905479
0.0053393	0.9902345	0.9903485
0.0043414	0.9893390	0.9893569
0.0035300	0.9887423	0.9889618
0.0028703	0.9880447	0.9880654
0.0023338	0.9881480	0.9881674
0.0018976	0.9875518	0.9877693
0.0015430	0.9875498	0.9876669

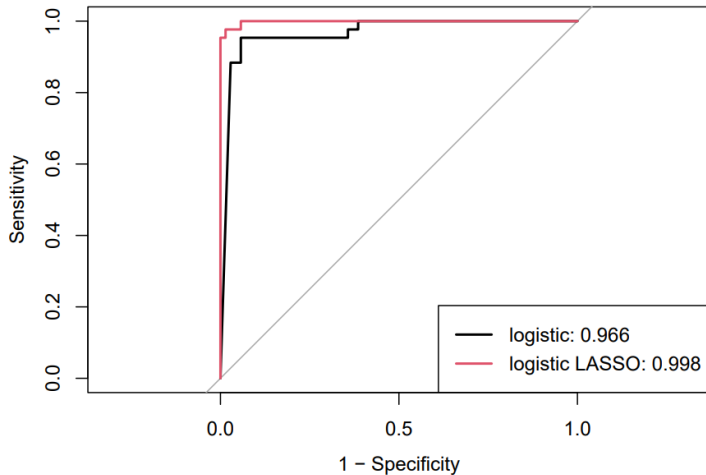
## Task 4 - Selecting predictors

predictor	ours_coef	cv.glmnet_coef
(Intercept)	-0.6138280	-0.6048659
radius_mean	0.0000000	0.0000000
texture_mean	0.3958698	0.3811372
perimeter_mean	0.0000000	0.0000000
area_mean	0.0000000	0.0000000
smoothness_mean	0.0000000	0.0000000
compactness_mean	0.0000000	0.0000000
concavity_mean	0.0000000	0.0000000
concave.points_mean	0.4850029	0.4476499
symmetry_mean	0.0000000	0.0000000
fractal_dimension_mean	0.0000000	0.0000000
radius_se	0.7660262	0.8379136
texture_se	0.0000000	0.0000000
perimeter_se	0.0000000	0.0000000
area_se	0.0000000	0.0000000
smoothness_se	0.0000000	0.0000000
compactness_se	0.0000000	0.0000000
concavity_se	0.0000000	0.0000000
concave.points_se	0.0000000	0.0000000
symmetry_se	0.0000000	0.0000000
fractal_dimension_se	-0.0364481	-0.0368622
radius_worst	2.5878169	2.5734608
texture_worst	0.4394404	0.4590245
perimeter_worst	0.0000000	0.0000000
area_worst	0.0000000	0.0000000
smoothness_worst	0.4980005	0.4945458
compactness_worst	0.0000000	0.0000000
concavity_worst	0.1437060	0.1253231
concave.points_worst	1.0087922	1.0692685
symmetry_worst	0.3617036	0.3641634
fractal_dimension_worst	0.0000000	0.0000000

## Task 4 - Compare 'optimal' and 'full' models - Table

Model	specificity	sensitivity	AUC
full	0.9268293	0.9305556	0.9661130
optimal	0.9545455	0.9855072	0.9983389

## Task 4 - Compare 'optimal' and 'full' models - Plot



# Discussions

- ▶ Lasso-logistic model performed better than the full logistic model.
- ▶ The most ideal performance is to classify each patient correctly, i.e.  $\text{Specificity} = \text{Sensitivity} = 1$
- ▶ Future work: Predictors need to be interpreted or have some important clinical significance.

## Reference

1. Hothorn et al. The design and analysis of benchmark experiments. Journal of Computational and Graphical Statistics (2005) vol. 14 (3) pp. 675-699
2. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33(1):1-22. PMID: 20808728; PMCID: PMC2929880.

# Q&A

- ▶ Thanks for listening!