Name ....

MTH210 (2024): Quiz 3

Roll. No. ....

For independent binary response Y_i, covariates $x_i \in \mathbb{R}^p$ for $i = 1, \dots, n$, and regression coefficient $\beta \in \mathbb{R}^p$, consider the following logistic regression model:

$$\Pr(Y_i = 1) = \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}}.$$

Goal is to obtain the ridge estimator of β. For a $\lambda \geq 0$, the ridge of β obtained is

$$\hat{\beta}_{\text{Ridge}} := \arg\min_{\beta \in \mathbb{R}^p} \left[\text{negative log likelihood} + \lambda \frac{\beta^T \beta}{2} \right]$$

7 1. Is the objective function convex? Explain mathematically why or why not?

3 2. Write down the steps of the Newton-Raphson algorithm to find $\hat{\beta}_{\text{Ridge}}$.

1. $\Pr(Y_i = 1) = \dfrac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}} =: p_i$ and $i = 1, \dots, n$ are independent

[following steps from notes) (should write all steps)

$$\log L(\beta \mid \underline{y}) = -\sum_{i=1}^n \log(1 + e^{x_i^T \beta}) + \sum_{i=1}^n y_i x_i^T \beta$$

-4 if negative not taken.

$$\Rightarrow nll := -\log L(\beta \mid y) = \sum_{i=1}^n \log(1 + e^{x_i^T \beta}) - \sum_{i=1}^n y_i x_i^T \beta$$

Now, the objective function is:

$$f(\beta) := nll + \lambda \frac{\beta^T \beta}{2} = \sum_{i=1}^n \log(1 + e^{x_i^T \beta}) - \sum_{i=1}^n y_i x_i^T \beta + \frac{\lambda}{2} \beta^T \beta$$

2

we wanto minimize this function. Let's take derivatives:

$$\nabla f(\beta) = -\sum_{i=1}^n x_i \left[y_i - \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}} \right] + \lambda \beta = -\sum_{i=1}^n x_i \left[y_i - \frac{1}{1 + e^{-x_i^T \beta}} \right] + \lambda \beta$$

3

$$\nabla^2 f(\beta) = \sum_{i=1}^n x_i \left[\frac{e^{-x_i^T \beta}}{(1 + e^{-x_i^T \beta})^2} \right] x_i^T + \lambda I_p$$

-2 if I_p not written

$$= \sum_{i=1}^n x_i \left[\frac{e^{x_i^T \beta}}{(1 + e^{x_i^T \beta})^2} \right] x_i^T + \lambda I_p = \boxed{X W X^T + \lambda I_p}$$

5

where $W =$ diagonal matrix of $\dfrac{e^{x_i^T \beta}}{(1 + e^{x_i^T \beta})^2}$ $i = 1, \dots, n$

A function f is convex if $\nabla^2 f(\beta)$ is positive-semi-definite.

Consider $a \in \mathbb{R}^p$ such that $a \neq 0$.

$$a^T \nabla^2 f(\beta) a = a^T (X^T W X + \lambda I_p) a$$ (-2) if error here.

$$= a^T (X^T W^{1/2} W^{1/2} X + \lambda I_p) a$$

$$= (W X a)^T (W X a) + \lambda a^T a$$

$$\geq 0.$$

7

Thus f is convex.

(2) NR Algorithm :

① Set $\beta_{(0)} = \underline{0}$ [(-1) if no starting value]

② For $k = 1, 2, \ldots\ldots$, set (-2) if not matrix inverse.

$$\beta_{(k+1)} = \beta_{(k)} - [\nabla^2 f(\beta_{(k)})]^{-1} [\nabla f(\beta_{(k)})]$$

③ Stop when $\|\beta_{(k+1)} - \beta_{(k)}\| < \varepsilon$ or when $\|\nabla f(\beta_{(k+1)})\| < \varepsilon$.