HW4(第一题)

葛宇泽

October 2021

1 Q1

$$
\begin{aligned}
l(\beta) &= \sum_{i=1}^{N}\left\{y_i \log p\left(x_i;\beta\right)+\left(1-y_i\right)\log\left(1-p\left(x_i;\beta\right)\right)\right\} \\
&= \sum_{i=1}^{N}\left\{y_i \log \frac{p\left(x_i;\beta\right)}{1-p\left(x_i;\beta\right)}+\log\left(1-p\left(x_i;\beta\right)\right)\right\} \\
\text{而：}\ & p\left(x_i;\beta\right)=P\left(Y=1 \mid X=x_i\right)=\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}} \\
& 1-p\left(x_i;\beta\right)=\frac{1}{1+e^{x_i'\beta}} \\
\text{代入得：}\ l(\beta) &= \sum_{i=1}^{N}\left\{y_i \log e^{x_i'\beta}-\log\left(1+e^{x_i'\beta}\right)\right\} \\
&= \sum_{i=1}^{N}\left\{y_i x_i^{\top}\beta-\log\left(1+e^{x_i'\beta}\right)\right]
\end{aligned}
$$

2 Q2

$$
\begin{aligned}
&\ell(\beta)=\sum_{i=1}^{N}\left[y_i x_i^{\top}\beta-\log\left(1+\exp\left(x_i^{\top}\beta\right)\right)\right] \\
&\frac{\partial l}{\partial \beta}=\sum_{i=1}^{n}\left[y_i-\frac{exp(x_i^{\top}\beta)}{1+\exp\left(x_i^{\top}\beta\right)}\right]x_i=\sum_{i=1}^{N}\left[y_i-p\left(x_i;\beta\right)\right]x_i
\end{aligned}
$$

In the formula:

$$
\begin{aligned}
&1-p\left(x_i;\beta\right)=\frac{1}{1+\exp\left(x_i^{\top}\beta\right)} \\
&-\frac{\partial p\left(x_i;\beta\right)}{\partial \beta_j}=-\frac{x_{ij}exp(x_i^{\top}\beta)}{1+\exp\left(x_i^{\top}\beta\right)}=-x_{ij}p\left(x_i;\beta\right)\left(1-p\left(x_i;\beta\right)\right) \\
&\frac{\partial p\left(x_i;\beta\right)}{\partial \beta}=x_i p\left(x_i;\beta\right)\left(1-p\left(x_i;\beta\right)\right)
\end{aligned}
$$

So:

$$
\begin{aligned}
\frac{\partial^2 l}{\partial \beta^\top \partial \beta}(k) &= -\sum_{i=1}^{N} \frac{\partial p\left(x_i;\beta\right)}{\partial \beta^\top} x_{ik} \\
&= -\sum_{i=1}^{n} x_i^\top p\left(x_i;\beta\right)\left(1-p\left(x_i;\beta\right)\right) x_{ik}
\end{aligned}
$$

Which is the k-th row of $\frac{\partial^2 l}{\partial \beta^\top \partial \beta}$

$$
\frac{\partial^2 l}{\partial \beta^\top \partial \beta} = -\sum_i p\left(x_i;\beta\right)\left(1-p\left(x_i;\beta\right)\right)\begin{bmatrix} x_i^\top x_{i1} \\ \vdots \\ x_i^\top x_{im} \end{bmatrix} = -\sum_i p\left(x_i;\beta\right)\left(1-p\left(x_i;\beta\right)\right) x_i x_i^\top
$$

m is the dimension of x_i

代码部分见 HW4_NR.html

3 Q3

$$
\begin{aligned}
& I\left(f\left(x^{+}\right)<f(x^{-})\right)=1-I\left(f\left(x^{+}\right)>f(x^{-})\right) \\
& l_{\text{rank }}=\frac{1}{m^{+} m^{-}} \sum_{x^{+} \in D^{+}} \sum_{x^{-} \in D^{-}}\left(I\left(f\left(x^{+}\right)<f(x^{-})\right)\right) \\
& =\frac{1}{m^{+} m^{-}} \sum_{x^{+} \in D^{+}} \sum_{x^{-} \in D^{-}}\left(1-I\left(f\left(x^{+}\right)>f\left(x^{-}\right)\right)\right) \\
& =1-\frac{1}{m^{+} m^{-}} \sum_{x^{+} \in D^{+}} \sum_{x^{-} \in D^{-}} I\left(f\left(x^{+}\right)>f(x^{-})\right) \\
& 1-l_{\text{rank }}=\frac{1}{m^{+} m^{-}} \sum_{x^{+} \in D^{+}} \sum_{x^{-} \in D^{-}} I\left(f\left(x^{+}\right)>f(x^{-})\right) \\
& =\sum_{x^{-} \in D^{-}} \frac{1}{m^{-}} \times \frac{1}{m^{+}} \sum_{x^{+} \in D^{+}} I\Big(f(x^{+})>f(x^{-})\Big)
\end{aligned}
$$

如图所示，我们将 ROC 曲线垂直于横轴分割为多个小长方形。在绘图过程中，如果选取的当前样本点为正例，则 ROC 曲线向上绘制 $1/m^{+}$ 而不向右绘制，故此类点与小长方形的个数无关。我们（由预测值从大到小）遍历每个负例，因为每个负例都会产生一个长为 $1/m^{-}$ 的小长方形。它的高为预测值比当前点更高的正例的个数 $\times 1/m^{+}$。（因为我们从预测值最高的点开始绘制，每遇到一个正例，就向上绘制制 $1/m^{+}$，直到当前点）将所有小长方形的面积相加，即为上式。

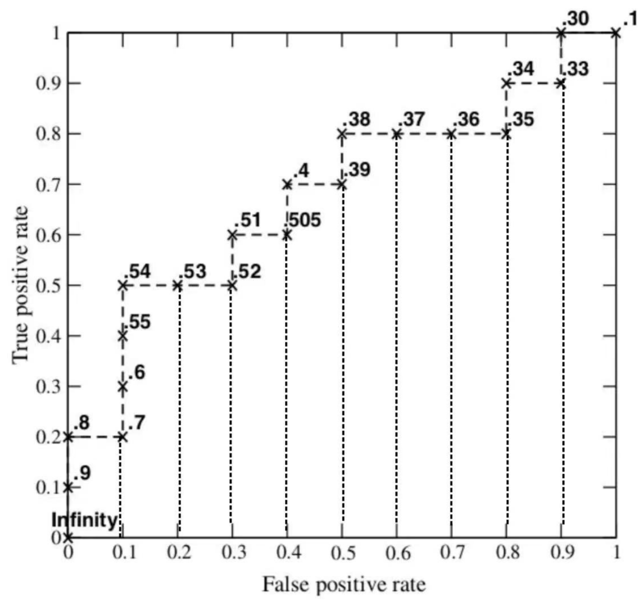

图 1: 将 ROC 曲线垂直于横轴分割为多个小长方体