Machine Learning Exercise Sheet 5

Linear Classification

Group_369

Fan Xue – fan98.xue@tum.de

Xing Zhou – xing.zhou@tum.de

Jianzhe Liu – jianzhe.liu@tum.de

November 24, 2021

Problem 3

a) Bernoulli distribution

b) x are classified as class 1 if $\mathrm{p}(y = 1 \mid x) > \mathrm{p}(y = 0 \mid x)$, which is equivalent to

$$\frac{\mathrm{p}(y = 1 \mid x)}{\mathrm{p}(y = 0 \mid x)} > 1$$

We take the logarithm of both sides and we get

$$\log \frac{\mathrm{p}(y = 1 \mid x)}{\mathrm{p}(y = 0 \mid x)} > 0$$

Simplify the left side

$$\begin{aligned}
\log \frac{\mathrm{p}(y = 1 \mid x)}{\mathrm{p}(y = 0 \mid x)} &= \log \frac{\mathrm{p}(x \mid y = 1)\mathrm{p}(y = 1)}{\mathrm{p}(x \mid y = 0)\mathrm{p}(y = 0)} \\
&= \log \frac{\mathrm{p}(x \mid y = 1)}{\mathrm{p}(x \mid y = 0)} \\
&= \log \frac{\lambda_1 \exp(-\lambda_1 x)}{\lambda_0 \exp(-\lambda_0 x)} \\
&= \log \frac{\lambda_1}{\lambda_0} - (\lambda_1 - \lambda_0)x
\end{aligned}$$

Consider both sides and simplify the inequality, we get

$$(\lambda_1 - \lambda_0)x < \log \lambda_1 - \log \lambda_0$$

Since $\lambda_1 - \lambda_0$ can be negative, we have

$$\begin{cases} x \in \left[0, \frac{\log \lambda_1 - \log \lambda_0}{\lambda_1 - \lambda_0}\right) & \text{if } \lambda_1 > \lambda_0 \\ x \in \left(\frac{\log \lambda_1 - \log \lambda_0}{\lambda_1 - \lambda_0}, \infty\right) & \text{otherwise.} \end{cases}$$

Problem 4

According to the given conditions in the question we can first have following loss function:

$$\begin{aligned} E(w) &= -\log p(\boldsymbol{y}|\boldsymbol{w}, \boldsymbol{X}) \\ &= -\sum_{i=1}^{N} y_i \log \sigma(\boldsymbol{w}^T \boldsymbol{x_i}) + (1 - y_i) \log(1 - \sigma(\boldsymbol{w}^T \boldsymbol{x_i})) \end{aligned}$$

where:

$$\sigma(a) = \left(\frac{1}{1 - e^{-a}}\right)$$

and:

$$\begin{cases} \boldsymbol{w}^T \boldsymbol{x_i} > 0, \text{if} \quad y_i = 1 \\ \boldsymbol{w}^T \boldsymbol{x_i} < 0, \text{if} \quad y_i = 0 \end{cases}$$

To get the optimized classification, we need to find the minimum of function $E(w)$.

Assuming that now we have $\boldsymbol{w} \to \infty$, then:

$$\begin{aligned} \underset{\boldsymbol{w}\to\infty}{E(w)} &= -\left(\sum_{\substack{i=1 \\ y_i=1}}^{N} \log \underset{\boldsymbol{w}\to\infty}{\sigma(\boldsymbol{w}^T \boldsymbol{x_i})} + \sum_{\substack{i=1 \\ y_i=0}}^{N} \log(1 - \underset{\boldsymbol{w}\to\infty}{\sigma(\boldsymbol{w}^T \boldsymbol{x_i})}) \right) \\ &= 0 \end{aligned}$$

This result clearly shows that the maximum likelihood parameter $\boldsymbol{w}$ of a logistic regression model has $||w|| \to \infty$, because only when $\boldsymbol{w} \to \infty$, can we achieve that minimum of $E(w)$.

What's more, if we want to get a $\boldsymbol{w}$ of finite magnitude, we can use weight regularization, for example we define $E_{new}(w) = E(w) + \boldsymbol{w}^T \boldsymbol{w}$.

In this way, it's obvious that when $\boldsymbol{w} \to \infty$, $E_{new}(w)$ is also infinite, then the optimized answer should be somewhere between 0 and infinite, thus finite.

Problem 5

The derivation is in follow:

$$\begin{aligned} \frac{e^{\boldsymbol{w}_1^T \boldsymbol{x}}}{e^{\boldsymbol{w}_1^T \boldsymbol{x}} + e^{\boldsymbol{w}_0^T \boldsymbol{x}}} &= \frac{1}{1 + e^{\boldsymbol{w}_0^T \boldsymbol{x} - \boldsymbol{w}_1^T \boldsymbol{x}}} \\ &= \frac{1}{1 + e^{-(\boldsymbol{w}_1 - \boldsymbol{w}_0)^T \boldsymbol{x}}} \\ &= \sigma\left((\boldsymbol{w}_1 - \boldsymbol{w}_0)^T \boldsymbol{x}\right) \end{aligned}$$

Problem 6

This question can be proved by following steps:

$$\begin{aligned}
\frac{\partial \sigma(a)}{\partial a} &= (-e^{-a})(-1)(1-e^{-a})^{-2} \\
&= \left(\frac{1}{1-e^{-a}}\right)\left(\frac{e^{-a}}{1-e^{-a}}\right) \\
&= \left(\frac{1}{1-e^{-a}}\right)\left(1-\frac{1}{1-e^{-a}}\right) \\
&= \sigma(a)(1-\sigma(a))
\end{aligned}$$

Problem 7

Let $\phi\left(x_1, x_2\right) = x_1 x_2$, we can observe that for all crosses $\phi\left(x_1, x_2\right) \leq 0$ and for all circles $\phi\left(x_1, x_2\right) \geq 0$, which means it is linearly separable. We can separate the crosses and circles with a single hyperplane $\phi\left(x_1, x_2\right) = 0$.

Problem 8

On the boundary Γ, the $\boldsymbol{x}$ must realize

$$p\left(y=1 \mid \boldsymbol{x}\right) = p\left(y=0 \mid \boldsymbol{x}\right)$$

It is equivalent to

$$\log \frac{p\left(y=1 \mid \boldsymbol{x}\right)}{p\left(y=0 \mid \boldsymbol{x}\right)} = 0$$

We expand

$$\begin{aligned}
\log \frac{p\left(y=1 \mid \boldsymbol{x}\right)}{p\left(y=0 \mid \boldsymbol{x}\right)} &= \frac{\frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma_1|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_1})^T \Sigma_1^{-1}(x-\mu_1)} \cdot \pi_1}{\frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma_0|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_0})^T \Sigma_0^{-1}(x-\mu_0)} \cdot \pi_0} \\
&= \frac{1}{2}\boldsymbol{x}^T\left(\Sigma_0^{-1}-\Sigma_1^{-1}\right)\boldsymbol{x} + \boldsymbol{x}^T\left(\Sigma_1^{-1}\boldsymbol{\mu_1} - \boldsymbol{\Sigma_0}^{-1}\boldsymbol{\mu}_0\right) \\
&\quad - \frac{1}{2}\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_0^T\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 + \log\frac{\pi_1}{\pi_0} + \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_0|}{|\boldsymbol{\Sigma}_1|} \\
&= \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c
\end{aligned}$$

where we define

$$\begin{aligned}
\boldsymbol{A} &= \frac{1}{2}\left(\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_1^{-1}\right) \\
\boldsymbol{b} &= \Sigma_1^{-1}\boldsymbol{\mu_1} - \boldsymbol{\Sigma_0}^{-1}\boldsymbol{\mu}_0 \\
c &= -\frac{1}{2}\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_0^T\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 + \log\frac{\pi_1}{\pi_0} + \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_0|}{|\boldsymbol{\Sigma}_1|}
\end{aligned}$$