ps2-2-solve

May 13, 2021

1 (a)

In logistic regression, we have:

$$
\begin{aligned}
h_\theta(x) &= \frac{1}{1+e^{-\theta^T x}} \\
L(\theta) &= \prod_{i=1}^{m} \left(h_\theta(x^{(i)})\right)^{y^{(i)}} \left(1-h_\theta(x^{(i)})\right)^{1-y^{(i)}} \\
\ell(\theta) &= \sum_{i=1}^{m} y^{(i)} \log h(x^{(i)}) + (1-y^{(i)}) \log(1-h(x^{(i)})) \\
\frac{\partial}{\partial \theta_j}\ell(\theta) &= \sum_{i=1}^{m} (y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)}
\end{aligned}
$$

because θ is the maximum likelihood parameters learned, so

$$
\begin{aligned}
\frac{\partial}{\partial \theta_j}\ell(\theta) &= 0 \\
\frac{\partial}{\partial \theta_0}\ell(\theta) &= 0 \\
\sum_{i=1}^{m} (y^{(i)} - h_\theta(x^{(i)}))x_0^{(i)} &= 0 \\
\sum_{i=1}^{m} (y^{(i)} - h_\theta(x^{(i)})) &= 0 \\
\sum_{i=1}^{m} h_\theta(x^{(i)}) &= \sum_{i=1}^{m} y^{(i)}
\end{aligned}
$$

let $(a, b) = (0, 1)$, then

$$
I_{a,b} = \{i | i \in \{1, \dots, m\}\} = \{1, \dots, m\}
$$

finally,

$$
\begin{gathered}
\sum_{i=1}^{m} h_\theta(x^{(i)}) = \sum_{i=1}^{m} y^{(i)} \\
\Sigma_{i \in I_{a,b}} h_\theta(x^{(i)}) = \Sigma_{i \in I_{a,b}} y^{(i)} \\
\frac{\Sigma_{i \in I_{a,b}} h_\theta(x^{(i)})}{|\{i \in I_{a,b}\}|} = \frac{\Sigma_{i \in I_{a,b}} y^{(i)}}{|\{i \in I_{a,b}\}|} \\
\frac{\Sigma_{i \in I_{a,b}} P(y^{(i)} = 1 | x^{(i)}; \theta)}{|\{i \in I_{a,b}\}|} = \frac{\Sigma_{i \in I_{a,b}} \mathbb{I}\{y^{(i)} = 1\}}{|\{i \in I_{a,b}\}|}
\end{gathered}
$$

2 (b)

No.

for example, let

$$
\begin{gathered}
h_\theta(x^{(1)}) = \frac{1}{2}, y^{(1)} = 1 \\
h_\theta(x^{(2)}) = \frac{1}{2}, y^{(2)} = 0 \\
h_\theta(x^{(i)}) = y^{(i)} \text{ for } i \in \{3, \ldots, m\}
\end{gathered}
$$

then, for any $(a, b) \subset [0, 1]$, we still have

$$
\Sigma_{i \in I_{a,b}} h_\theta(x^{(i)}) = \Sigma_{i \in I_{a,b}} y^{(i)}
$$

But, obviously, our model dosen't achieve perfect accuracy.

No.

model has perfect accuracy means

$$
1 > h_\theta(x^{(i)}) > \frac{1}{2} \text{ for } i \in \{i | i \in \{1, \ldots, m\}, y^{(i)} = 1\}
$$

let $(a, b) = (0.5, 1]$, then

$$
\Sigma_{i \in I_{a,b}} y^{(i)} = 1 > \Sigma_{i \in I_{a,b}} h_\theta(x^{(i)})
$$

Obviously our model isn't perfectly calibrated.

3 (c)

we want to maximize log-likelihood $\ell(\theta)$

$$\ell(\theta) = \sum_{i=1}^{m} y^{(i)} \log h(x^{(i)}) + (1 - y^{(i)}) \log(1 - h(x^{(i)}))$$

it means we want to minimize loss J

$$J = -\ell(\theta)$$

including $ L_2 $ regularization in our loss$ J_{L2}$

$$J_{L2} = J + \frac{1}{2}\lambda\|\theta\|_2^2 = -\ell(\theta) + \frac{1}{2}\lambda\|\theta\|_2^2$$

let

$$\begin{aligned} \frac{\partial}{\partial\theta_0} J_{L2} &= 0 \\ \frac{\partial}{\partial\theta_0}\big(-\ell(\theta) + \frac{1}{2}\lambda\|\theta\|_2^2\big) &= 0 \\ \lambda\theta_0 - \sum_{i=1}^{m}(y^{(i)} - h_\theta(x^{(i)})) &= 0 \\ \lambda\theta_0 + \sum_{i=1}^{m} h_\theta(x^{(i)}) &= \sum_{i=1}^{m} y^{(i)} \end{aligned}$$

if $\theta \neq 0$, then property not hold.

[ ]: