Exercise

01

TUM Department of Informatics

Supervised by Prof. Dr. Stephan Günnemann

Informatics 3 - Professorship of Data Mining and Analytics

Submitted by Marcel Bruckner (03674122)

Julian Hohenadel (03673879)

Kevin Bein (03707775)

Submission date Munich, November 13, 2019

Linear Classification

Problem 1:

a) It must be Bernoulli because it is a binary classification model. y can only become 0 or 1.

b) For a x to be classified as 1 the probability of $p(y=1|x)$ must be greater than the probability of $p(y=0|x)$ which is equal to $\log \frac{p(y=1|x)}{p(y=0|x)} > 0$.

$$\begin{aligned}
\log \frac{p(y=1|x)}{p(y=0|x)} &= \log \frac{p(x|y=1) \cdot p(y=1)}{p(x|y=0) \cdot p(y=0)} \\
&= \log \frac{p(x|y=1) \cdot \frac{1}{2}}{p(x|y=0) \cdot \frac{1}{2}} \\
&= \log \frac{p(x|y=1)}{p(x|y=0)} \\
&= \log \frac{\lambda_1 \exp(-\lambda_1 x)}{\lambda_0 \exp(-\lambda_0 x)} \\
&= \log \frac{\lambda_1}{\lambda_0} + \lambda_0 x - \lambda_1 x \\
&= (\log(\lambda_1) - \log(\lambda_0)) + (\lambda_0 - \lambda_1)x \\
&\Rightarrow (\lambda_0 - \lambda_1)x > \log(\lambda_0) - \log(\lambda_1) \\
&\Rightarrow x > \frac{\log(\lambda_0) - \log(\lambda_1)}{\lambda_0 - \lambda_1}
\end{aligned}$$

For $\lambda_0 > \lambda_1$, x must be bigger than $\frac{\log(\lambda_0)-\log(\lambda_1)}{\lambda_0-\lambda_1}$. In the other case where $\lambda_1 > \lambda_0$, x must be bigger or equal to 0 but smaller than $\frac{\log(\lambda_0)-\log(\lambda_1)}{\lambda_0-\lambda_1}$.

Problem 2:

$$\begin{aligned}
\hat{w} &= \arg\min_w E(w) \\
&= \arg\min_w -\prod_{i=1}^{N} \sigma(w^T x_i)^{y_i} (1 - \sigma(w^T x_i))^{1-y_i} \\
&= \arg\max_w \prod_{i=1}^{N} \sigma(w^T x_i)^{y_i} (1 - \sigma(w^T x_i))^{1-y_i}
\end{aligned}$$

For $w \longrightarrow \infty$, $\sigma(w^T x^i) \longrightarrow 0$. It follows that $\arg\max_w E(w) = \prod_{i=1}^{N} 0^{y_i} \cdot (1-0)^{1-y_i} = 0$.

Ultimately, this will lead to a heaviside step function which will classify every training point with a posterior probability of $p(y|x) = 1$. This can be prevented bei either testing against an upper threshold or penalizing large weights (as already sugegsted in the lecture: $E(w) = -\ln p(y - |w, X) + \lambda \, ||w||_q^2$.

Problem 3:

Softmax $\sigma_{\text{soft}}(x) = \frac{\exp(x_i)}{\sum_{i=1}^{C} \exp(x_k)}$ and Sigmoid $\sigma_{\text{sig}} = \frac{1}{1+\exp(-a)}$. For $C = \{1, 2\}$ we show that $\sigma_{\text{soft}} \stackrel{!}{=} \sigma_{\text{sig}}$.

$$
\begin{aligned}
\sigma_{\text{soft}}(a) &= \frac{\exp(w_1^T x)}{\sum_{i=1}^{C} \exp(w_i^T x)} \\
&= \frac{\exp(w_1^T x)}{\exp(w_1^T x) + \exp(w_2^T x)} \\
&= \frac{1}{1 + \exp(w_2^T x) / \exp(w_1^T x)} \\
&= \frac{1}{1 + \exp(w_2^T x - w_1^T x)} \\
&= \frac{1}{1 + \exp(-(w_1 - w_2)^T x)} \\
&= \sigma_{\text{sig}}(\bar{w}^T x) \\
&= \sigma_{\text{sig}}(a)
\end{aligned}
$$

Problem 4:

Similar to to the lecture, we need to map to a different space. For example, simply determining in which quadrant a point is located in, is sufficient.

$$
\phi(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 < 0, x_2 > 0 \text{ or } x_1 > 0, x_2 < 0 \\ 1 & \text{else} \end{cases}
$$

(0 if x is the top right or bottom left quadrant and 1 otherwise)

Appendix

We confirm that the submitted solution is original work and was written by us without further assistance. Appropriate credit has been given where reference has been made to the work of others.

Munich, November 13, 2019, Signature Marcel Bruckner (03674122)

Munich, November 13, 2019, Signature Julian Hohenadel (03673879)

Munich, November 13, 2019, Signature Kevin Bein (03707775)