Assignment 3

Devansh Jain, 190100044

17th Oct 2021

Contents

1 Logistic Regression

1.1 CS 337: Logistic Regression

Given Soft-max expression and Categorical cross-entropy function of Multi-class Logistic Regression, we substitute $K = 2$.

$$\begin{aligned}
P(Y=1|\mathbf{w}_1,\phi(\mathbf{x})) &= \frac{e^{\mathbf{w}_1^T\phi(\mathbf{x})}}{e^{\mathbf{w}_1^T\phi(\mathbf{x})}+e^{\mathbf{w}_2^T\phi(\mathbf{x})}} = \frac{1}{1+e^{(\mathbf{w}_2^T-\mathbf{w}_1^T)\phi(\mathbf{x})}}\\
P(Y=2|\mathbf{w}_2,\phi(\mathbf{x})) &= \frac{e^{\mathbf{w}_2^T\phi(\mathbf{x})}}{e^{\mathbf{w}_1^T\phi(\mathbf{x})}+e^{\mathbf{w}_2^T\phi(\mathbf{x})}} = \frac{e^{(\mathbf{w}_2^T-\mathbf{w}_1^T)\phi(\mathbf{x})}}{1+e^{(\mathbf{w}_2^T-\mathbf{w}_1^T)\phi(\mathbf{x})}}\\
E(\mathbf{w}_1,\mathbf{w}_2) &= -\frac{1}{N}\sum_{i=1}^{N} y_1^{(i)}\log(P(Y=1|\mathbf{w}_1,\phi(\mathbf{x}^{(i)}))) + y_2^{(i)}\log(P(Y=2|\mathbf{w}_2,\phi(\mathbf{x}^{(i)})))\\
&= -\frac{1}{N}\sum_{i=1}^{N}\frac{y_1^{(i)}+y_2^{(i)}e^{(\mathbf{w}_2^T-\mathbf{w}_1^T)\phi(\mathbf{x})}}{1+e^{(\mathbf{w}_2^T-\mathbf{w}_1^T)\phi(\mathbf{x})}}
\end{aligned}$$

Now, we substitute the one-hot vector $[y_1^{(i)}, y_2^{(i)}]$ with $[y^{(i)}, 1-y^{(i)}]$ and $\mathbf{w}_1 - \mathbf{w}_2$ with $\mathbf{w}$.

$$\begin{aligned}
P(Y=1|\mathbf{w},\phi(\mathbf{x})) &= \frac{1}{1+e^{-\mathbf{w}^T\phi(\mathbf{x})}} = \sigma_{\mathbf{w}}(\mathbf{x})\\
P(Y=2|\mathbf{w},\phi(\mathbf{x})) &= \frac{e^{-\mathbf{w}^T\phi(\mathbf{x})}}{1+e^{-\mathbf{w}^T\phi(\mathbf{x})}} = 1-\sigma_{\mathbf{w}}(\mathbf{x})\\
E(\mathbf{w}) &= -\frac{1}{N}\sum_{i=1}^{N}\frac{y^{(i)}+(1-y^{(i)})e^{-\mathbf{w}^T\phi(\mathbf{x})}}{1+e^{-\mathbf{w}^T\phi(\mathbf{x})}}\\
&= -\frac{1}{N}\sum_{i=1}^{N}\frac{y^{(i)}}{1+e^{-\mathbf{w}^T\phi(\mathbf{x})}} + \frac{(1-y^{(i)})e^{-\mathbf{w}^T\phi(\mathbf{x})}}{1+e^{-\mathbf{w}^T\phi(\mathbf{x})}}\\
&= -\frac{1}{N}\sum_{i=1}^{N} y^{(i)}\sigma_{\mathbf{w}}(\mathbf{x}) + (1-y^{(i)})(1-\sigma_{\mathbf{w}}(\mathbf{x}))
\end{aligned}$$

Hence, cross entropy function used to train a binary logistic regression is a special case of the categorical cross entropy function given above.

1.2 CS 337: Logistic Regressions Decision surface

The model assigns a value of $y = +1$ to a point $\mathbf{x}$ if $P(y=+1|\mathbf{x}) = \dfrac{1}{1+\exp(-\mathbf{w}^T\mathbf{x})} \geq \theta$ which is equivalent to $\mathbf{w}^T\mathbf{x} \geq -\log\left(\dfrac{1-\theta}{\theta}\right)$ and this forms the decision boundary.
When we say that a model is linear, we mean that its predictions are a linear function of its parameters.
Logistic regression is considered as a linear model because the **decision boundary used for classification purpose is linear** (i.e. $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$).