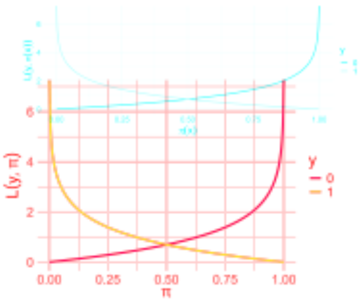
Introduction to Machine Learning

Advanced Risk Minimization

Logistic regression (Deep-Dive)



Learning goals

- Derive the gradient of the logistic regression
- Derive the Hessian of the logistic regression
- Show that the logistic regression is a convex problem

LOGISTIC REGRESSION: GRADIENT / 2

$$= \sum_{i=1}^{n} \left(\pi\left(\mathbf{x}^{(i)} \mid \boldsymbol{\theta}\right) - y^{(i)}\right) \left(\mathbf{x}^{(i)}\right)^\top .$$

$$= (\pi(\mathbf{X}| \boldsymbol{\theta}) - \mathbf{y})^\top \mathbf{X}$$

where $\mathbf{X} = \left(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\right)^\top \in \mathbb{R}^{n \times d}, \mathbf{y} = \left(y^{(1)}, \dots, y^{(n)}\right)^\top$,
$\pi(\mathbf{X}| \boldsymbol{\theta}) = \left(\pi\left(\mathbf{x}^{(1)} \mid \boldsymbol{\theta}\right), \dots, \pi\left(\mathbf{x}^{(n)} \mid \boldsymbol{\theta}\right)\right)^\top \in \mathbb{R}^n$.

$\Rightarrow$ The gradient $\nabla_{\boldsymbol{\theta}} \mathcal{R}_{\text{emp}} = \left(\frac{\partial}{\partial \boldsymbol{\theta}} \mathcal{R}_{\text{emp}}\right)^\top = \mathbf{X}^\top (\pi(\mathbf{X}| \boldsymbol{\theta}) - \mathbf{y})$

This formula can now be used in gradient descent and its friends.

LOGISTIC REGRESSION: CONVEXITY

Finally, we check that logistic regression is a convex problem:

We define the diagonal matrix $\bar{\mathbf{D}} \in \mathbb{R}^{n \times n}$ with diagonal

$$\left(\sqrt{\pi\left(\mathbf{x}^{(1)} \mid \boldsymbol{\theta}\right))(1 - \pi\left(\mathbf{x}^{(1)} \mid \boldsymbol{\theta}\right)}, \dots, \sqrt{\pi\left(\mathbf{x}^{(n)} \mid \boldsymbol{\theta}\right)\left(1 - \pi\left(\mathbf{x}^{(n)} \mid \boldsymbol{\theta}\right)}\right)$$

which is possible since π maps into $(0, 1)$.

With this, we get for any $\mathbf{w} \in \mathbb{R}^d$ that

$$\mathbf{w}^\top \nabla^2_{\boldsymbol{\theta}} \mathcal{R}_{\text{emp}} \mathbf{w} = \mathbf{w}^\top \mathbf{X}^\top \bar{\mathbf{D}}^\top \bar{\mathbf{D}} \mathbf{X} \mathbf{w} = (\bar{\mathbf{D}} \mathbf{X} \mathbf{w})^\top \bar{\mathbf{D}} \mathbf{X} \mathbf{w} = \|\bar{\mathbf{D}} \mathbf{X} \mathbf{w}\|_2^2 \geq 0$$

since obviously $\mathbf{D} = \bar{\mathbf{D}}^\top \bar{\mathbf{D}}$.

$\Rightarrow \nabla^2_{\boldsymbol{\theta}} \mathcal{R}_{\text{emp}}$ is positive semi-definite $\Rightarrow \mathcal{R}_{\text{emp}}$ is convex.