Differential Geometry

Instructor: Jianfeng Lin

Notes Taker: Zejin Lin

TSINGHUA UNIVERSITY.

linzj23@mails.tsinghua.edu.cn

lzjmaths.github.io

February 28, 2025

Contents

1 Regression

$$\min_{\omega\in\mathbb{R}^m} \frac{1}{2N}\|\Phi\omega - y\|^2 + \lambda C(\omega) \tag{1.1}$$

Lasso: $C = \|\omega\|_1$. **Ridge regression**: $C = \|\omega\|_2$.

subgradient of f:

$$\partial f(x_0) = \{g | f(x) \geqslant f(x_0) + g^T(x - x_0)\}$$

In particular,

$$\partial|x| = \begin{cases} 1, & x > 0 \\ -1, & x < 0 \\ [-1, 1], & x = 0 \end{cases}$$

1.1 Binary classification problem

one-hot encoding for the output $\{\binom{1}{0}, \binom{0}{1}\}$. It can be understood as the probability for each class and can take continuous values.

A **linear hypothesis space** is $\{u(x) : u = \omega^T x, x \in \mathbb{R}^n, \omega \in \mathbb{R}^n\}$.

Softmax:Map the extracted feature u to the space of one-hot codes

$$\mu = \frac{1}{1+e^{-u}}, \quad 1 - \mu = \frac{e^{-u}}{1+e^{-u}} = \frac{1}{1+e^{u}}$$

$$KL(p, q) = \int p(\log p - \log q) \tag{1.2}$$

For p real probability, to minimize (1.2), suffices to minimize

$$-\int p \log q_\theta \mathrm{d}x = -\sum_{x_i} \log q_\theta(x_i)$$

which is called **Maximum likelihood (cross entropy)**

$$-\sum \log p(y_i|x_i, \omega) = \sum -y_i \log \mu_i - (1 - y_i)\log(1 - \mu_i)$$

We reduce to minimize the thing above.

1.2 Gradient Descent

$$J(\theta) = \sum_{i=1}^{N} L(f_\theta(x_i), y_i), \quad \theta^{t+1} = \theta^t - \eta_t \frac{\partial J(\theta)}{\partial \theta}|_{\theta=\theta^t}$$

Index

List of Theorems