

Regressão Linear

Martin Lobe

Orientador: Luiz Rafael dos Santos

Introdução

Em regressão, buscamos uma função f da qual mapeia pontos em \mathbb{R}^D para valores correspondentes em $f(x) \in \mathbb{R}$. O objetivo é criar um modelo que a partir da informação de treinamento, consiga generalizar o suficiente para prever os resultados de novos pontos de entrada.

1 Revisão de Conceitos

Definição 1 (Verossimilhança). A verossimilhança é um valor que corresponde a quão provável que nossos dados gerados tenham vindo de um modelo linear. Isso quer dizer, nossos dados que serão obtidos através da própria função possuem uma maior verossimilhança. Alguns modelos podem ser menos adequados, e isso mostra uma verossimilhança baixa.

Definição 2 (I.I.D). Significa independentemente, igualmente distribuído.

- Verossimilhança
- i.i.d.?

2 Formulando o problema

Abordaremos o problema de forma probabilística. O motivo para isso é devido ao *Observation Noise*. *Observation Noise* é a variabilidade do erro que possui em relação

aos valores observados e as funções que estão por trás dessas medidas feitas. Estamos considerando que essas variabilidades são igualmente e independentemente distribuídas

$$p(y | x) = \mathcal{N}(y | f(x), \sigma^2)$$

(Estamos comparando a aproximação de p com a função gaussiana sendo \mathcal{N})

Na equação mostrada, temos que $x \in \mathbb{R}^D$, os nossos valores recebidos, enquanto $y \in \mathbb{R}$ é o valor resultado pela função nesse ponto, sendo o nosso objetivo.

$$y = f(x) + \epsilon; \epsilon \in \mathcal{N}(0, \sigma^2)$$

Como podemos ver abaixo, f é a função que estamos querendo descobrir, com o valor de y sendo próximo dele. Essa proximidade é uma probabilidade dada pela função Gaussiana. Essa função f é desconhecida, e o objetivo é encontrar uma função que seja próximo o suficiente dela.

Vamos começar com funções paramétricas. Começaremos considerando que a variância σ^2 é conhecida, e trabalharemos em encontrar um parâmetro θ . Em regressão linear, teremos o caso especial em que θ aparece linearmente no modelo.

$$p(y | x) = \mathcal{N}(y | x^T \theta, \sigma^2) \Leftrightarrow y = x^T \theta$$

Exemplo 1 (Exemplo). Para $x, \theta \in \mathbb{R}$, o modelo de regressão linear se torna uma reta, ou seja uma função linear, o parâmetro θ é a inclinação dessa reta.

Teorema 3 (Theorema de Gauss Markov). *Para um modelo linear padrão, o menor estimador de variância linear sem tendência para β_i é dado pelo i -ésimo componente de $\hat{\beta}_i$*

Proposição 4 (Prop. 1). *Se $x \in \mathbb{R}^{m \times n}$ e $\text{rank}(x) = n$ então $(x^\top x)$ é inversível (não-singular).*

Proposição 5 (Prop. 2). *As equações normais sempre tem solução, ou equivalentemente, $x^\top \beta \in \text{Im}(x^\top x) \subset \mathbb{R}^n$.*

Lema 6. $\text{Im}(x^\top x) = \text{Im}(x^\top)$.