

$y = X\beta + \epsilon$, X 的设置, ϵ 的意义

$\text{logit}(\pi) = ?$

$\pi = \text{sigmoid}(\cdot)$

等价于线性拟合 $\text{logit}(\pi) = x^T \beta$

MLE 的连乘形式, log 求和形式, 求导推导

Classification 的情况下, y_i 取值正负 1, $p(y_i|x_i) = ?$

三类 models 的基本定义, 建模哪一类概率分布

Covariance matrix

Softmax function

Loss function, typically $\text{loss} = \sum_i -\log p(y_i|x_i)$

Logistic loss 【是不是源于 logistic regression】

Exponential loss

Hinge loss (SVM 中)

Zero-one loss (不能求导)

最小二乘回归的解

Entropy 的定义

crossEntropy 的定义

KL 散度的定义

MLE $\Rightarrow$ 降低数据分布于模型对应分布之间的 KL 散度

梯度下降公式

SGD

动量的概念和公式

了解 discriminative model: learning from errors

Descriptive model: learning from dream

Langevin: 知道在梯度上加噪声可以防止 local minimum

LDA: 目标函数 + 结论

SVM

Margin 的定义, margin 值大小与分类正确的关系

物理意义: 到分类面的距离

W = 支持向量的线性和

SVM 的目标函数 (两种形式)

为什么要惩罚 $\|w\|^2$

哪些样本上 $\alpha_i > 0$ 的, $=0$ 的

$W^T x_i + b$ 写成核函数的形式 $\Rightarrow$ 只要得到 $K(x_i, x_j)$ 就可以得到分类值

核函数的条件: 对称性, 半正定

经典核函数: RBF kernel

Regularized learning

Overfitting

Ridge regression 公式，解，物理意义：防止极少量特征维度主导分类

Kernel regression 公式，解

$f(x)$ =核函数表达下的公式

Kernel regression 本质是 Ridge regression 在核函数空间的对应

DNN 中应用 - 》weight decay

Spline regression 的定义

LASSO regression, 公式，解??，物理意义：学习稀疏的特征。

shrinkage (by subtracting λ) • selection (via thresholding at λ)

Primal form & Dual form

Bayesian regression

证明加入 ridge loss 以后的解空间

Neural networks

会设计一个浅层的神经网络（三层 sigmoid 网络）

CNN 中：kernel size 的设置，padding 的设置

输出输出张量 size 的计算，参数量，计算量的多少

设计一个简单的 CNN：注意设计 CNN 时的一些【坑】

反向传播的公式：能写出公式，而且写出公式中每个矩阵是什么样的

crossEntropy 的梯度写进反向传播，sigmoid 触发函数写进反向传播

batch normalization 的公式

残差网络的基本形式

LSTM, GRU 不需要背，但是需要可以知道每个 gate function 的意义

GAN 的公式

大概知道本质在优化 JSD 分布

VAE 基本公式

ELBO 有现实的优化目标

Lecture 6

PCA 的定义与计算

如何用 PCA 进行降维

t-SNE 的优化目标

LLE 的优化目标

EM 算法：在特殊分布下，给出根本目标的公式，优化目标 Q 的公式

EM 框架下如何计算对 theta 的梯度