IML Summary

Lasse Fierz - lfierz

Version: 2. Februar 2023

Basics

- General p-norm: $||x||_p = (\sum_{i=1}^{n} |x_i|^p)^{1-p}$
- Taylor: $f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \mathbb{O}(x^3)$
- Power series of exp.: $exp(x) := \sum_{k=0}^{\infty} \frac{x^k}{k!}$
- Entropy: H(X) = $\mathbb{E}_X \left[-log\mathbb{P}(X=x)\right]$
- KL-Divergence: $D_{KL}(P||Q) = \sum_{x \in \mathbb{X}} P(x) log\left(\frac{P(x)}{Q(x)}\right) \geq 0$
- $1 - z \leq exp(-z)$
- Cauchy-Schwarz: $|\mathbb{E}\left[X,Y\right]|^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$
- Jensens Inequality: for a convex f(X): $f(\mathbb{E}(X)) \leq \mathbb{E}(f(X))$

Probability Theory:

- Gaussian: $\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2})$
- $(N)(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}|}} exp(-\frac{1}{2}(\boldsymbol{x}-\mu)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\mu))$
- $X \sim \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}), Y = A + BX \Rightarrow Y \sim \mathcal{N}(A + B\boldsymbol{\mu}, B\boldsymbol{\Sigma^{-1}}B^T)$
- Binomial Distr.: $f(k,j;p) = \mathbb{P}(X=x) = \binom{n}{k} p^k(1-p)^{n-k}$
- $\mathbb{V}(X) = \mathbb{E}\left[(X - \mathbb{E}(X))^2\right] = \mathbb{E}(X^2) - \left[\mathbb{E}(X)\right]^2$
- $\mathbb{V}\left[X+Y\right] = \mathbb{V}\left[X\right] + \mathbb{V}\left[Y\right] + 2Cov(X,Y)$
- $Cov(X,Y) = \mathbb{E}\left[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right]$
- $Cov(aX,bY) = abCov(X,Y)$

Calculus

- $\int uv'dx = uv - \int u'vdx$ • $\frac{\partial}{\partial x}\frac{g}{h} = \frac{g'h}{h^2} - \frac{gh'}{h^2}$
- $\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{b^T A x}) = A^T b$ • $\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{b^T x}) = \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x^T b}) = b$
- $\frac{\partial}{\partial \boldsymbol{X}}(\boldsymbol{c^T X^T b}) = \boldsymbol{b}\boldsymbol{c}^T$ • $\frac{\partial}{\partial \boldsymbol{X}}(\boldsymbol{c^T X b}) = \boldsymbol{c b^T}$
- $\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x^T A x}) = (\boldsymbol{A^T} + \boldsymbol{A})\boldsymbol{x} \overset{\text{A sym.}}{=} 2\boldsymbol{Ax}$
- $\frac{\partial}{\partial \boldsymbol{X}}Tr(\boldsymbol{X^T A}) = A$ • Tr.trick: $\boldsymbol{x^T A x} \overset{\text{inner prod.}}{=} TR(\boldsymbol{x^T A x}) \overset{\text{cyclic perm.}}{=} Tr(\boldsymbol{x x^T A}) = Tr(\boldsymbol{A x x^T})$
- $|X^{-1}| = |X|^{-1}$ • $\frac{\partial}{\partial \boldsymbol{X}}log|\boldsymbol{X}| = \boldsymbol{X}^{-T}$ • $\frac{\partial}{\partial x}|x| = \frac{x}{|x|}$
- $\frac{\partial}{\partial \boldsymbol{x}}||\boldsymbol{x}||_2 = \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x^T x}) = 2\boldsymbol{x}$
- $\frac{\partial}{\partial \boldsymbol{x}}||\boldsymbol{x} - \boldsymbol{b}||_2 = \frac{\boldsymbol{x}-\boldsymbol{b}}{||\boldsymbol{x}-\boldsymbol{b}||_2}$
- $\frac{\partial}{\partial \boldsymbol{x}}||\boldsymbol{x}||_1 = sgn(\boldsymbol{x})$
- $\sigma(x) = \frac{1}{1+exp(-x)} \Rightarrow$
- $\nabla\sigma(x) = \sigma(x)(1-\sigma(x)) = \sigma(x)\sigma(-x)$
- $tanhx = \frac{2sinhx}{2coshx} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
- $\nabla tanhx = 1 - tanh^2x$

(Linear) Regression

General Regression: find $\hat{y} = f(x) \leftrightarrow \underset{\hat{y}(x)}{min}||y - \hat{y}(x)||_2^2$.

Linear Regression: Weights are applied linearly:
$f(x) = \beta x$ or nonlinear **base fct**: $f(x) = \beta\phi(x)$
Multidim.: $\underset{\beta}{min}||\boldsymbol{Y} - \boldsymbol{X\beta}||^2$, $\boldsymbol{Y} \in \mathbb{R}^n, \boldsymbol{X} \in \mathbb{R}^{nxd}, \boldsymbol{\beta} \in \mathbb{R}^d$

If

SVM

Gradient Descent and Convexity

Model Selection

Classification

Kernels

Neural Networks

Clustering

Dimensionality Reduction

Statistical Perspective

Generative Modelling

Gaussian Mixture Model

Additionals