

07-09-ArtificialIntelligence

Created on 20190819.

Last modified on 2022 年 6 月 5 日.

目录

Chapter 1 自动推理

Chapter 2 Computer Vision

Computer Graphics, from image to model. 机器视觉，需要猜测，包括理解、意义。

Computer Vision, from model(modeling, simulation) to image, 计算机图形学，不需要猜测。

2.1 Introduction

寒武纪物种爆发，因为眼睛的出现 computer vision begins with Larry Robert,1963. the edges define the shape, 1966 mit AI group established. David Marr book < vision > 视觉是简单的形状开始的，视觉是分层的视觉识别模型：1) Generalized Cylinder 简单物体组成 Brooks 1979; 2) Pictorial Structure, 基础元素之间用弹簧连接。语义分割：Normalized Cut, shi 1997 viola jones face detector detection , pascal imagenet, sigmoid 换成 relu, 使得 2012 比赛冠军, 和 LeCun1998 年的手写数字没有太大区别

2.2 Others

2.2.1 图片分类

最近邻算法：L1 distance , 曼哈顿距离，

$$d_1 = \sum |I_1^i - I_2^i| \quad (2.1)$$

特点：每一个测试图片的计算时间是恒定的；Flann 是实例库 L2 distance, l1 距离的平方根 K nearest neighbor, k 是超参数 train data 分子集比较超参数, cross-validation 线性分类器，需要识别纹理，对无纹理的识别困难 the threshold at zero $\max(0, -)$ function is often called the **hinge loss**.

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + \Delta)$$

初始 w 为 0 时，损失等于类数量-1 **regularization loss**

$$L = \underbrace{\frac{1}{N} \sum_i L_i}_{\text{data loss}} + \underbrace{\lambda R(W)}_{\text{regularization loss}} \quad (2.2)$$

$$L = \frac{1}{N} \sum_i \sum_{j \neq y_i} [\max(0, f(x_i; W)_j - f(x_i; W)_{y_i} + \Delta)] + \lambda \sum_k \sum_l W_{k,l}^2 \quad (2.3)$$

weight regularization: L2: $\sum W^2$

L1: $\sum |W|$

elastic net $\sum \alpha W^2 + |W|$ max norm regularization dropout

Chapter 3 Computer Graphics

表 3.1: Usage of Computer Graphics

Name	Notes
Video games	全局光照，越亮游戏画面越好
movies	特效。远离现实的较容易。日常见的想要特效看得真实较为困难
animations	zootopia 中的动物头发。
Design	Autodesk.Ikea, render
Visualization	science, engineering, medicine, journalism, etc.
VR, AR.	
Digital Illustration.	
Simulation	dust bowl, black hole.
Graphical User Interfaces.	
Topography.	The quick brown fox jumps over the lazy dog.

physics of lighting and shading

representing and operating shapes in 3D.

3.1 Rasterization

Project geometry primitives(3D triangles or polygons) into fragments(pixels).

3.2 Curves and Meshes

3.3 Ray Tracing

Shoot rays from the camera though each pixel.

3.4 Animation and Simulation

Chapter 4 语音及多媒体信息处理

Chapter 5 机器学习

得到的的 DF

5.1 ANN, 人工神经网络

的 55 得到的的

ANN 的实质是通过已知的两个空间的一对子空间，寻找两个空间的映射关系，希望通过局部性质对整体有所刻画。

5.1.1 基础 1-凸优化

概念

极大似然估计

$$\begin{aligned} \min -L(\mu, \sigma) \\ \sigma \geq 0 \end{aligned} \tag{5.1}$$

最小二乘

$$\min f_0(x) = \|\mathbf{Ax} - \mathbf{b}\|^2 \tag{5.2}$$

where $\mathbf{A}_{n \times k}$, $\mathbf{b} \in \mathbb{R}^n$, for $\mathbf{x} \in \mathbb{R}^k$.

凸优化局部最优 = 全局最优

$$\begin{aligned} \min f_0(x) \\ f_i(x) \leq b_i, i = 1, \dots, m. \end{aligned} \tag{5.3}$$

$$\begin{aligned}
& \forall x_1, x_2 \in \Omega, \lambda \in (0, 1) \\
& \text{convex set} : \lambda x_1 + (1 - \lambda)x_2 \in \Omega \\
& \text{convex function} : f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)
\end{aligned} \tag{5.4}$$

上境图是凸集 = 凸函数

凸组合 (重心):

$$S = w^i x_i, \text{ where } \sum w^i = 1 \tag{5.5}$$

凸包: x_i 的全部凸组合

凸闭包: f 的凸闭包的上境图, 是 f 的上境图的凸包。

Jensen 不等式: 对于凸函数 f , 有:

$$\sum w_i f(x_i) \geq f\left(\sum w_i x_i\right) \tag{5.6}$$

大部分不等式来自于 $x^2 \geq 0$, 或者 Jensen 不等式, 如:

$$\left. \begin{aligned} f &= -\ln(x) \\ w_i &\equiv \frac{1}{n} \end{aligned} \right\} \Rightarrow \left(\prod_{i=1}^n a_i \right)^{\frac{1}{n}} \leq \frac{\sum_{i=1}^n a_i}{n} \tag{5.7}$$

$$\left. \begin{aligned} f &= x^2 \\ x_i &= \frac{a_i}{b_i} \\ w_i &= \frac{b_i^2}{\sum b_i^2} \end{aligned} \right\} \Rightarrow \sum a_i^2 \sum b_i^2 \geq \left(\sum a_i b_i \right)^2 \tag{5.8}$$

性质

凸集性质:

凸集交集是凸集;

凸集的线性映射是凸集;

平行光源投影 (到任意平面上) 保持凸集;

点光源投影 (集合所有元素都除以同一个元素) 保持凸集, $\Omega_{\hat{n}} = \{x_i/x_n | x_i \in \Omega\}$;

点光源投影 (椎体) 保持凸集, $\{tx_i | x_i \in \Omega\}$;

凸集合边界可微, 则边界切平面是凸集的支撑平面;

凸集边界二阶可微, 则边界点曲率向量指向集合内部, 曲率向量是加速度方向或受力方向;

凸函数性质:

固定凸函数某些变量仍然是凸函数;

凸函数的非负线性组合是凸函数;

凸函数一阶可微, 则一阶近似不大于函数本身, $f(x) \geq f(x_0) + (\nabla f(x_0))^T \cdot (x - x_0)$;

凸函数二阶可微, 则 Hessian 阵半正定;

凸函数 $f(x_1, \dots, x_n)$, 有 $g(x_i) = \inf$ 是凸函数;

升维锥体保持凸性 $f: \mathbb{R}^n \mapsto \mathbb{R} \Rightarrow g(x, t) = tf(x/t): \mathbb{R}^{n+1} \mapsto \mathbb{R}$;

凸集分离定理:

\mathbb{R}^n 中两不相交非空凸集 C 和 D, 存在 $a \in \mathbb{R}^n, b \in \mathbb{R}$, 使得 $a^T x_C \leq b$ & $a^T x_D \geq b$, 几何意义是两个凸集在超平面 $a^T x = b$ 两侧, 其中 a 是超平面的法向量。超平面是 n 维空间中的 n-1 维平面。

5.1.2 基础 2-对偶问题

共轭函数

任意函数 f 的共轭函数: $f^*(y) = \sup(y^T x - f(x))$, 右边括号里是勒让德变换, 相当于在找从函数到超平面 $y^T x$ 的距离最大值, 函数返回从曲率等于超平面的 f(x) 沿着 y 的方向到超平面的最大值。

性质: f^* 是凸函数;

若 g 是 f 的凸闭包, $g^* = f^*$;

f 是凸函数时有 $f^{**} = f$;

$f(x) + f^*(y) \geq x^T y$;

f 是凸函数可微时, $f^*(y) = \nabla f(x)x - f(x)$;

$g(x) = f(Ax + b) \Rightarrow g^*(y) = f^*(A^{-T}y) - b^T A^{-T}y$;

$f(u, v) = f_1(u) + f_2(v) \Rightarrow f^*(w, z) = f_1^*(w) + f_2^*(z)$;

如:

$$\begin{aligned}
 f(x) &= x \ln x \Rightarrow \\
 f^*(y) &= \sup_x (yx - x \ln x) \\
 \frac{d(yx - x \ln x)}{dx} &= 0 \Rightarrow x^* = e^{y-1} \\
 \therefore f^*(y) &= ye^{y-1} - e^{y-1}(y-1) = e^{y-1}
 \end{aligned} \tag{5.9}$$

拉格朗日对偶函数

对于 \mathbb{R} 上的优化问题:

$$\begin{aligned} \min f_0(x) \\ f_i(x) &\leq 0 \\ h_i(x) &= 0 \end{aligned} \quad (5.10)$$

优化点 x^* , 最优值 p^* ;

拉格朗日量 $\mathbb{R}^{n+m+p} \mapsto \mathbb{R}$

$$L(x, \lambda, v) = f_0(x) + \sum \lambda_i f_i(x) + \sum v_i h_i(x) \quad (5.11)$$

取 L 的下确界, 定义拉格朗日对偶函数

$$g(\lambda, v) = \inf_x L(x, \lambda, v) \quad (5.12)$$

对于 $\lambda \geq 0$ 有 $g(\lambda, v) \leq p^*$, 对偶函数能提供下界, 因此希望最大化 g 。对偶问题的最大值点 (λ^*, v^*) , 最大值 d^* 。

例子 1:

$$\begin{aligned} \min c^T x \\ \text{s.t. : } \begin{cases} x_i \geq 0 \Rightarrow -x_i \leq 0 \\ A^T x = b \end{cases} \\ \Rightarrow L = C^T x - \lambda^T x + v^T (A^T x - b) \\ = (C^T - \lambda^T + v^T A^T) x - v^T b \\ g(\lambda, v) = \begin{cases} -\infty \\ -v^T b, C^T - \lambda^T + v^T A^T = 0 \end{cases} \\ \therefore \min v^T b \\ \text{s.t. : } \lambda \geq 0 \\ C^T - \lambda^T + v^T A^T = 0 \end{aligned} \quad (5.13)$$

限制条件是线性时:

$$\begin{aligned}
 g &= \inf_x [f_0(x) + \lambda^T(Ax - b) + v^T(Cx - d)] \\
 &= -b^T\lambda - d^Tv + \inf_x [(A^T\lambda + C^Tv)^T x + f_0(x)] \\
 &= -b^T\lambda - d^Tv - f^*(-a^t\lambda - c^tv)
 \end{aligned} \tag{5.14}$$

例子 2, 最小化向量范数:

$$\min |x|, \text{ where } Ax = b \tag{5.15}$$

$$f^*(y) = \sup_x (\lambda^T x - |x|) = \begin{cases} 0, & |y| \leq 1 \\ +\infty, & |y| > 1 \end{cases} \tag{5.16}$$

$$\begin{aligned}
 g &= \inf_x [|x| + \lambda^T(Ax - b)] \\
 &= -b^T\lambda + \sup_x (-A^T x - |x|) \\
 &\therefore \max -b^T\lambda \\
 &|A^T\lambda| \leq 1
 \end{aligned} \tag{5.17}$$

例子 3, 最大熵:

$$\begin{aligned}
 &\max -\sum x_i \ln x_i \\
 &Ax \leq b \\
 &1^T x = 1
 \end{aligned} \tag{5.18}$$

$$\begin{aligned}
 y^* &= \sum e^{y_i - 1} \Rightarrow \\
 g &= \inf_x \left[\sum x_i \ln x_i + \lambda^T(Ax - b) + v^T(x - 1) \right] \\
 &= -b^T\lambda - v + \sup_x \left(-\lambda A^T x - vx - \sum x_i \ln x_i \right) \\
 &= -b^T\lambda - v - \sum e^{-\lambda^T A - v^T - 1} \\
 &\therefore \max g \\
 &\lambda \geq 0
 \end{aligned} \tag{5.19}$$

对偶性

弱对偶性: $d^* \leq p^*$

强对偶性: $d^* = p^*$

slater 条件, 对于凸优化问题, 如果存在取到不等号的点, 就满足强对偶性条。如线性规划、最小

二乘、最大熵问题都满足。

凸优化求解 (KKT)

$$\begin{aligned}
 f_i(x^*) &\leq 0 \\
 h_i(x^*) &= 0 \\
 \lambda_i^* &\geq 0 \\
 \lambda_i^* f_i(x^*) &= 0 \\
 \nabla_x L(x^*, \lambda^*, v^*) &= 0
 \end{aligned} \tag{5.20}$$

例子 1 kkt 求解优化问题:

$$\begin{aligned}
 \min \quad & \frac{1}{2} x^T p x + q^T x + r \\
 \text{Ax} &= b \\
 \because h_i(x^*) &= 0 \\
 \nabla L &= 0 \\
 \therefore Ax^* &= b \\
 Px + q + A^T v^* &= 0 \\
 \Rightarrow \begin{bmatrix} A & 0 \\ p & A^T \end{bmatrix} \begin{bmatrix} x^* \\ v^* \end{bmatrix} &= \begin{bmatrix} b \\ 0 \end{bmatrix}
 \end{aligned} \tag{5.21}$$

支持向量机 SVM

$$\begin{aligned}
 a^T x_c &> b \ \& \ a^T x_d < d \\
 a^T x_c - b &\geq t \ \& \ a^T x_d - d \leq -t
 \end{aligned} \tag{5.22}$$

在这里 a 是单位向量, 保证与原问题等价, 转化为 $|a| \leq 1$:

$$\begin{aligned}
 \min \quad & -t \\
 -a^T x_c + b &\leq -t \\
 a^T x_d - b &\leq -t \\
 -t &\leq 0 \\
 |a|^2 &\leq 1
 \end{aligned} \tag{5.23}$$

5.1.3 基础 3-数学

微积分

梯度写作列向量, Hessian matrix:

$$\begin{aligned}\nabla f(\mathbf{x}) &= \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \\ \mathbf{H}(\mathbf{x}) &= \nabla^2 f(\mathbf{x}) = \left[\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} \right]\end{aligned}\quad (5.24)$$

一阶导数为 0, 可能是极值点, 同时二阶导数为 0 的时候就是鞍点 saddle point, 判断鞍点可用三阶导数。矢量的泰勒级数展开

$$f(\mathbf{x}_i + \delta) \approx f(\mathbf{x}_i) + \nabla^T f(\mathbf{x}_i)\delta + \frac{1}{2}\delta^T \nabla^2 f(\mathbf{x}_i)\delta \quad (5.25)$$

为什么梯度方向是上升最快的方向, 因为泰勒公式中可以看到, 取梯度方向时, 向量共线, 夹角 0, 模最大; 同样有负梯度方向最小。通常而言, 方向更重要, 步长没有方向那么重要
把二次项也考虑进来, 就叫牛顿法

概率论部分

累积分布函数 $F(x) = P(x \leq x_0)$

概率密度函数 $f(x) = \frac{d}{dx}F(x)$ 高斯分布: 独立同分布收敛于高斯分布, 加三四项就类似高斯了

$$f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (5.26)$$

贝叶斯公式:

$$\begin{aligned}P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ P(A|B)P(B) &= P(B|A)P(A) \\ f(x|y) &= \frac{f(x, y)}{f(y)}\end{aligned}\quad (5.27)$$

贝叶斯分类, 分词, 图像识别, 邮件过滤。

5.1.4 基础 4-Regression

线性回归

目标函数 f 写成 x 带偏置的线性函数，有

$$f(\mathbf{x}) = \theta^T \mathbf{x} \quad (5.28)$$

损失函数，构造 convex 的，如

$$J = \text{ave}(\hat{\mathbf{y}} - \mathbf{y})^2 \quad (5.29)$$

梯度下降

$$\theta^{(i+1)} = \theta^{(i)} - \alpha \nabla J \quad (5.30)$$

逻辑回归

目标函数加一层阶跃函数， $\text{sign}(f)$ ，sigmoid 函数

$$\frac{1}{1 + e^{-x}} \quad (5.31)$$

损失函数，目的是对于判断错误的能很明显放大错误

$$C_{\text{oss}} = \begin{cases} -\log(\hat{\mathbf{y}}), \mathbf{y} = 1 \\ -\log(1 - \hat{\mathbf{y}}), \mathbf{y} = 0 \end{cases} \quad (5.32)$$

防止过拟合，添加权值作为正则化项

$$\begin{aligned} J &= -\text{ave} [\mathbf{y} \cdot \log(\hat{\mathbf{y}}) + 1 - \mathbf{y} \cdot \log(1 - \hat{\mathbf{y}})] + \lambda \text{ave} |\theta|^2 \\ \theta_i &= \theta_i = \alpha \frac{\partial J}{\partial \theta_j} \end{aligned} \quad (5.33)$$

多分类问题：one-vs-rest, 得到每个点属于每个类的概率。损失函数如 linearSVM (可用 SGD 求解)，或者交叉熵

$$\begin{aligned} L_i &= \sum_{j \neq y_i} \max [0, f(x_i, w)_j - f(x_i, w)_{y_i} + \Delta] \\ &= \sum_{j \neq y_i} \max (0, w_j^T x_i - w_{y_i}^T x_i + \Delta) \\ L &= \text{ave}(L_i) + \lambda \sum_k \sum_l \end{aligned} \quad (5.34)$$

交叉熵，指数函数保持正，归一化，指数函数为防止过大，用常数 C 做平滑：

$$L = - \sum y_i \log \frac{e^{f_{y_i}}}{\sum e^{f_{y_i}}} = - \sum y_i \log \frac{e^{f_{y_i} + C}}{\sum e^{f_{y_i} + C}} \quad (5.35)$$

vision.stanford.edu/teaching/cs231n

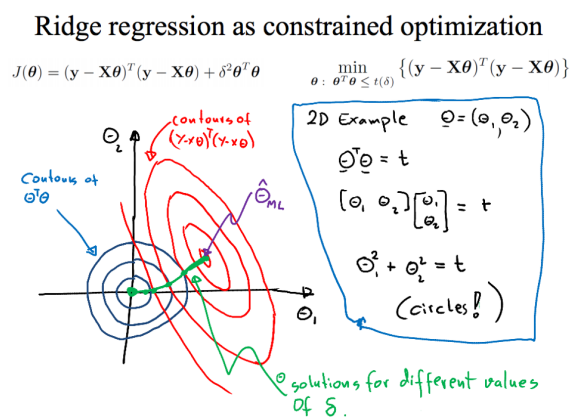


Figure 5.1: 正则项的几何意义

需要查找?? HIFT,JIST,HOG

5.1.5 基础概念

停止准则：与真值误差小于预设；两次迭代差小于预设；达到预设迭代次数

5.1.6 分类

单层感知器

单层感知器（Perceptron）解决线性可分问题，在高维空间用一个超平面划分样本。Rosenblatt 证明两类模式线性可分时算法收敛。

$$Y = \text{sgn} \left([w_i, b] [x_i, 1]^T \right) = \text{sgn}(\mathbf{W}^T \mathbf{X}) \quad (5.36)$$

$$\mathbf{W}_{n+1} = \mathbf{W}_n + \eta (\mathbf{Y}_{\text{real}} - \mathbf{Y}_n) \mathbf{X}_n$$

二值化，分类的边界距离某一类很接近。常采用纠错学习规则的学习算法，把偏置作为固定输入
局限性：不能解决线性不可分问题。奇异样本训练时间长。只适合单层。

5.2 计算智能

5.2.1 遗传算法的理论和应用

5.2.2 免疫模型与算法的基本原理及其应用

5.2.3 人工神经网络的理论和应用

5.2.4 网络智能信息检索与数据挖掘

5.3 增强学习

5.4 深度学习

Chapter 6 计算机语言学、机器翻译 及自然语言理解

自然语言处理

Chapter 7 智能信息处理

Chapter 8 仿脑计算理论与仿脑计算机

Chapter 9 虚拟现实，增强现实，虚拟环境

Chapter 10 专家系统

Chapter 11 机器人