Chapter 1 计算机视觉、语音及多媒体信息处理

D

Chapter 2 计算机语言学、机器翻译及自然语言理解

Chapter 3 仿脑计算理论与仿脑计算机

Chapter 4 智能信息处理

Chapter 5 虚拟现实，增强现实，虚拟环境

Chapter 6 机器学习

得到的的 DF

6.1 ANN, 人工神经网络

的 55 得到的的的

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。

6.1.1 基础 1-凸优化

概念

极大似然估计

$$\begin{aligned} &min - L\left(\mu, \sigma\right) \\ &\sigma \geqslant 0 \end{aligned} \tag{6.1}$$

最小二乘

$$min f_0(x) = \left|\mathbf{A}\mathbf{x} - \mathbf{b}\right|^2 \tag{6.2}$$

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

$$\begin{aligned} &min f_0(x) \\ &f_i(x) \leqslant b_i, i = 1, \ldots, m. \end{aligned} \tag{6.3}$$

$$\begin{aligned}
&\forall x_1, x_2 \in \Omega, \lambda \in (0,1) \\
&convex\ set: \lambda x_1 + (1-\lambda)x_2 \in \Omega \\
&convex\ function: f\left(\lambda x_1 + (1-\lambda)x_2\right) \leqslant \lambda f(x_1) + (1-\lambda) f(x_2)
\end{aligned} \tag{6.4}$$

上镜图是凸集 = 凸函数

凸组合 (重心):

$$S = w^i x_i, where \sum w^i = 1 \tag{6.5}$$

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：

$$\sum w_i f\left(x_i\right) \geqslant f\left(\sum w_i x_i\right) \tag{6.6}$$

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

$$\left.\begin{aligned} f &= -ln(x) \\ w_i &\equiv \frac{1}{n} \end{aligned}\right\} \Rightarrow \left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}} \leqslant \frac{\sum\limits_{i=1}^{n} a_i}{n} \tag{6.7}$$

$$\left.\begin{aligned} f &= x^2 \\ x_i &= \frac{a_i}{b_i} \\ w_i &= \frac{b_i^2}{\sum b_i^2} \end{aligned}\right\} \Rightarrow \sum a_i^2 \sum b_i^2 \geqslant \left(\sum a_i b_i\right)^2 \tag{6.8}$$

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，$f(x) \geqslant f(x_0) + (\nabla f(x_0))^T \cdot (x - x_0)$;
凸函数二阶可微，则 Henssen 阵半正定；
凸函数 $f(x_1, ..., x_n)$，有 $g(x_i) = inff$ 是凸函数；
升维椎体保持凸性 $f : \mathbb{R}^n \mapsto \mathbb{R} \Rightarrow g(x, t) = tf(x/t) : \mathbb{R}^{n+1} \mapsto \mathbb{R}$；

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 $a^T x_C \leqslant b$ & $a^T x_D \geqslant b$，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。

6.1.2 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;
$f(x) + f^*(y) \geqslant x^T y$;
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:

$$\begin{aligned} & f(x) = xlnx \Rightarrow \\ & f^*(y) = \sup_x (yx - xlnx) \\ & \frac{d(yx - xlnx)}{dx} = 0 \Rightarrow x^* = e^{y-1} \\ & \therefore f^*(y) = ye^{y-1} - e^{y-1}(y - 1) = e^{y-1} \end{aligned} \tag{6.9}$$

拉格朗日对偶函数

对于 $\mathbb{R}$ 上的优化问题：

$$
\begin{aligned}
& minf_0(x) \\
& f_i(x) \leqslant 0 \\
& h_i(x) = 0
\end{aligned} \tag{6.10}
$$

优化点 x*，最优值 p*；
拉格朗日量 $\mathbb{R}^{n+m+p} \mapsto \mathbb{R}$

$$
L(x,\lambda,\upsilon) = f_0(x) + \sum \lambda_i f_i(x) + \sum \upsilon_i h_i(x) \tag{6.11}
$$

取 L 的下确界，定义拉格朗日对偶函数

$$
g(\lambda,\upsilon) = \inf_x L(x,\lambda,\upsilon) \tag{6.12}
$$

对于 $\lambda \geqslant 0$ 有 $g(\lambda,\upsilon) \leqslant p^*$，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^*,υ^*)，最大值 d*。

例子 1：

$$
\begin{aligned}
& minc^T x \\
& s.t.: \begin{cases} x_i \geqslant 0 \Rightarrow -x_i \leqslant 0 \\ A^T x = b \end{cases} \\
\Rightarrow L &= C^T x - \lambda^T x + \upsilon^T (A^T x - b) \\
&= \left(C^T - \lambda^T + \upsilon^T A^T\right) x - \upsilon^T b \\
& g(\lambda,\upsilon) = \begin{cases} -\infty \\ -\upsilon^T b, C^T - \lambda^T + \upsilon^T A^T = 0 \end{cases} \\
& \therefore \min \upsilon^T b \\
& s.t.: \lambda \geqslant 0 \\
& C^T - \lambda^T + \upsilon^T A^T = 0
\end{aligned} \tag{6.13}
$$

限制条件是线性时：

$$\begin{aligned} g &= \inf_x \left[f_0(x) + \lambda^T(Ax-b) + \upsilon^T(Cx-d)\right] \\ &= -b^T\lambda - d^T\upsilon + \inf_x \left[\left(A^T\lambda + C^T\upsilon\right)^T x + f_0(x)\right] \\ &= -b^T\lambda - d^T\upsilon - f^*\left(-a^t\lambda - c^t\upsilon\right) \end{aligned} \tag{6.14}$$

例子 2，最小化向量范数：

$$in|x|, where\ Ax = b \tag{6.15}$$

$$f^*(y) = \sup_x \left(\lambda^T x - |x|\right) = \begin{cases} 0, |y| \leqslant 1 \\ +\infty, |y| > 1 \end{cases} \tag{6.16}$$

$$\begin{aligned} &g = \inf_x \left[|x| + \lambda^T(Ax-b)\right] \\ &= -b^T\lambda + \sup_x \left(-A^T x - |x|\right) \\ &\therefore max - b^T\lambda \\ &|A^T\lambda| \leqslant 1 \end{aligned} \tag{6.17}$$

例子 3，最大熵：

$$\begin{aligned} &max - \sum x_i ln x_i \\ &Ax \leqslant b \\ &1^T x = 1 \end{aligned} \tag{6.18}$$

$$\begin{aligned} &y^* = \sum e^{y_i - 1} \Rightarrow \\ &g = \inf_x \left[\sum x_i ln x_i + \lambda^T(Ax-b) + \upsilon^T(x-1)\right] \\ &= -b^T\lambda - \upsilon + \sup_x \left(-\lambda A^T x - \upsilon x - \sum x_i ln x_i\right) \\ &= -b^T\lambda - \upsilon - \sum e^{-\lambda^T A - \upsilon^T - 1} \\ &\therefore maxg \\ &\lambda \geqslant 0 \end{aligned} \tag{6.19}$$

对偶性

弱对偶性：$d^* \leqslant p^*$

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）

$$
\begin{aligned}
&f_i(x^*) \leqslant 0 \\
&h_i(x^*) = 0 \\
&\lambda_i^* \geqslant 0 \\
&\lambda_i^* f_i(x^*) = 0 \\
&\nabla_x L(x^*, \lambda^*, \upsilon^*) = 0
\end{aligned} \tag{6.20}
$$

例子 1 kkt 求解优化问题：

$$
\begin{aligned}
&min\ \frac{1}{2} x^T p x + q^T x + r \\
&Ax = b \\
&\because h_i(x^*) = 0 \\
&\nabla L = 0 \\
&\therefore Ax^* = b \\
&Px + q + A^T \upsilon^* = 0 \\
&\Rightarrow \begin{bmatrix} A & 0 \\ p & A^T \end{bmatrix} \begin{bmatrix} x^* \\ \upsilon^* \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}
\end{aligned} \tag{6.21}
$$

支持向量机 SVM

$$
\begin{aligned}
&a^T x_c > b \ \&\ a^T x_d < d \\
&a^T x_c - b \geqslant t \ \&\ a^T x_d - d \leqslant -t
\end{aligned} \tag{6.22}
$$

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

$$
\begin{aligned}
&min - t \\
&- a^T x_c + b \leqslant -t \\
&a^T x_d - b \leqslant -t \\
&- t \leqslant 0 \\
&|a|^2 \leqslant 1
\end{aligned} \tag{6.23}
$$

6.1.3 基础 3-数学

微积分

梯度写作列向量，Hessian matrix:

$$\begin{aligned}\nabla f(\mathbf{x}) &= \frac{\partial f(\mathbf{x})}{\partial x} \\ \mathbf{H}(\mathbf{x}) = \nabla^2 f(\mathbf{x}) &= \left[\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}\right]\end{aligned} \tag{6.24}$$

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

$$f(\mathbf{x}_i + \delta) \approx f(\mathbf{x}_i) + \nabla^T f(\mathbf{x}_i)\delta + \frac{1}{2}\delta^T \nabla^2 f(\mathbf{x}_i)\delta \tag{6.25}$$

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 $F(x) = P(x \leqslant x_0)$
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

$$f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{6.26}$$

贝叶斯公式：

$$\begin{aligned}P(A|B) &= \frac{P(A \bigcap B)}{P(B)} \\ P(A|B)P(B) &= P(B|A)P(A) \\ f(x|y) &= \frac{f(x,y)}{f(y)}\end{aligned} \tag{6.27}$$

贝叶斯分类，分词，图像识别，邮件过滤。

6.1.4 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

$$f(\mathbf{x}) = \theta^T \mathbf{x} \tag{6.28}$$

损失函数，构造 convex 的，如

$$J = ave(\hat{\mathbf{y}} - \mathbf{y})^2 \tag{6.29}$$

梯度下降

$$\theta^{(i+1)} = \theta^{(i+1)} - \alpha \nabla J \tag{6.30}$$

逻辑回归

目标函数加一层阶跃函数，$sign(f)$，sigmoid 函数

$$\frac{1}{1+e^{-x}} \tag{6.31}$$

损失函数，目的是对于判断错误的能很明显放大错误

$$Coss = \begin{cases} -log(\hat{\mathbf{y}}), \mathbf{y} = 1 \\ -log(1-\hat{\mathbf{y}}), \mathbf{y} = 0 \end{cases} \tag{6.32}$$

防止过拟合，添加权值作为正则化项

$$\begin{aligned} J &= -ave\left[\mathbf{y}. \cdot log(\hat{\mathbf{y}}) + 1 - \mathbf{y}. \cdot log(1-\hat{\mathbf{y}})\right] + \lambda ave|\theta|^2 \\ \theta_i &= \theta_i = \alpha \frac{\partial J}{\partial \theta_j} \end{aligned} \tag{6.33}$$

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

$$\begin{aligned} L_i &= \sum_{j \neq y_i} \max\left[0, f(x_i, w)_j - f(x_i, w)_{y_i} + \Delta\right] \\ &= \sum_{j \neq y_i} \max\left(0, w_j^T x_i - w_{y_i}^T x_i + \Delta\right) \\ L &= ave(L_i) + \lambda \sum_k \sum_l \end{aligned} \tag{6.34}$$

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：

$$L = -\sum y_i \log \frac{e^{f_{y_i}}}{\sum e^{f_{y_i}}} = -\sum y_i \log \frac{e^{f_{y_i}+C}}{\sum e^{f_{y_i}+C}} \tag{6.35}$$

vision.stanford.edu/teaching/cs231n

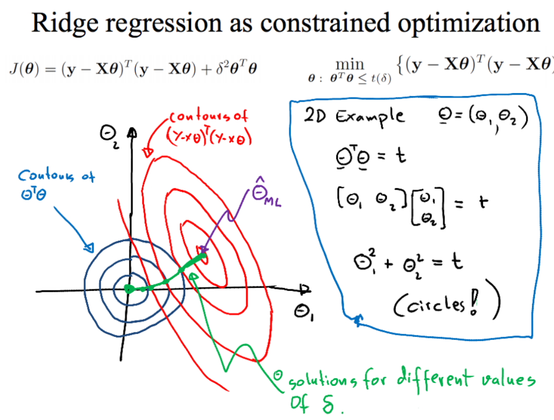

Figure 6.1: 正则项的几何意义

需要查找?? HIFT,JIST,HOG

6.1.5 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

6.1.6 分类

单层感知器

单层感知器（Perception）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。

$$\begin{aligned} \boldsymbol{Y} &= \operatorname{sgn}\left([w_i, b][x_i, 1]^T\right) = \operatorname{sgn}(\boldsymbol{W}^T \boldsymbol{X}) \\ \boldsymbol{W}_{n+1} &= \boldsymbol{W}_n + \eta(\boldsymbol{Y}_{\text{real}} - \boldsymbol{Y}_n)\boldsymbol{X}_n \end{aligned} \tag{6.36}$$

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。

6.2 计算智能

6.2.1 遗传算法的理论和应用

6.2.2 免疫模型与算法的基本原理及其应用

6.2.3 人工神经网络的理论和应用

6.2.4 网络智能信息检索与数据挖掘

6.3 增强学习

6.4 深度学习