决策树

Error = noise + bias + variance

集成学习 $\quad X \Rightarrow \frac{1}{N}\sum_{i=1}^{N} x_i$

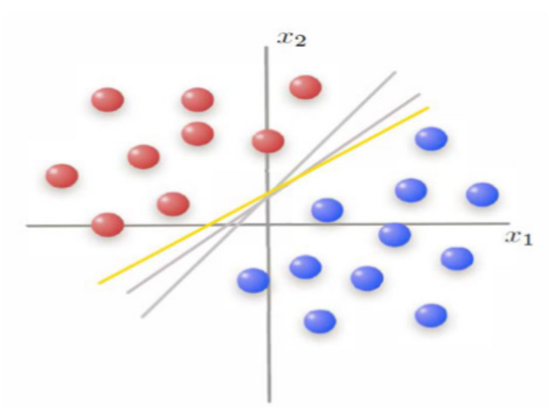
$$E((X-E(X))^2) = \sigma^2$$

$$E\left(\frac{1}{N}\sum X_i - E(X)\right)^2$$
$$= \frac{1}{N^2}E\left(\sum X_i - NE(X)\right)^2$$
$$= \frac{1}{N^2}E\left(X_1-E(X)+X_2-E(X)\cdots\right)^2$$
$$= \frac{1}{N^2}E\Big[(X_1-E(X))^2+\cdots(X_N-E(X))^2$$
$$+\sum_{i\neq j}(X_i-E(X))(X_j-E(X))\Big]$$
$$= \frac{1}{N^2}\left[E(X_1-E(X))^2+\cdots E(X_N-E(X))^2\right]$$
$$= \frac{1}{N^2}N\sigma^2 = \frac{\sigma^2}{N}$$

方差下降

SVM (max margin classifier)

Support Vector Machine



评判标准:
分界线离最近的点越远越好

Is there an "optimal" way to separate the data?
这么多分类界面，哪个是最优的呢？

Basic problem:

Given N data point, $x_1, \cdots x_N \in R^m$
each with a label $d_i = \pm 1$
find a hyperplane

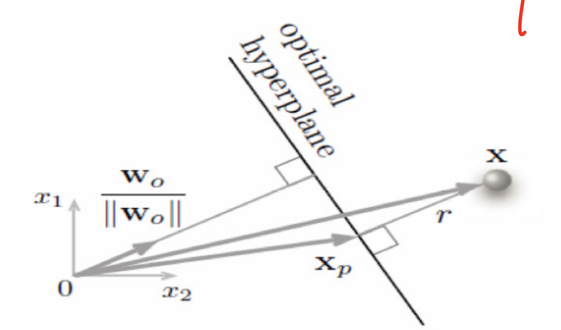
$$w^T x + b = 0 \quad (w \in R^m, b \in R)$$

that separates data into two groups

$$\boxed{\begin{cases} g(\mathbf{x}_i) = \mathbf{w}_o^T \mathbf{x}_i + b_o \geq +1 & \text{for } d_i = +1 \\ g(\mathbf{x}_i) = \mathbf{w}_o^T \mathbf{x}_i + b_o \leq -1 & \text{for } d_i = -1 \end{cases}}$$

$$\Rightarrow d_i(w_o^T x_i + b_o) \geq 1$$

同时要最大化间隔 r, $\begin{cases} r \propto \frac{1}{\|w\|} & \text{(证明见ppt)} \\ w_o \perp \text{hyperplane} \end{cases}$



$\Rightarrow$

优化模型 Find w 和 b

$$\begin{cases} \min: \Phi(w) = \frac{1}{2} w^T w \\ \text{Subject to}: d_i(w^T x_i + b) \geq 1 \end{cases}$$

接下来用KKT条件化成对偶问题

KKT条件（拉格朗日乘子推广） https://zhuanlan.zhihu.com/p/38163970

①等式约束优化

$$\min \ f(x)$$
$$s.t. \ g(x) = 0$$

$$\Rightarrow \ L(x,\lambda) = f(x) + \lambda g(x)$$

$$\Rightarrow \ \min L(x,\lambda)$$

$$\begin{cases} \nabla_x L = \nabla f + \lambda \Delta g = 0 \\ \nabla_\lambda L = g(x) = 0 \end{cases} \Rightarrow \text{求解即得 } x^*$$

②不等式约束优化

$$\min \ f(x)$$
$$s.t. \ g(x) \le 0$$

$$\Rightarrow \ L(x,\lambda) = f(x) + \lambda g(x)$$

$$\Rightarrow \begin{cases} \nabla_x L = \nabla f + \lambda \nabla g = 0 \\ g(x) \le 0 \\ \lambda \ge 0 \\ \lambda g(x) = 0 \end{cases}$$

$$\min \ f(x)$$
$$s.t. \ g(x) \ge 0$$

$$\Rightarrow \ L(x,\lambda) = f(x) - \lambda g(x)$$

$$\Rightarrow \begin{cases} \nabla_x f = \nabla f - \lambda \nabla g(x) = 0 \\ g(x) \ge 0 \end{cases}$$

$$\begin{cases} \lambda \geq 0 \\ \lambda g(x) = 0 \end{cases}$$

接下来我们将约束等式 $g(\mathbf{x})=0$ 推广为不等式 $g(\mathbf{x}) \leq 0$ 。考虑这个问题

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & g(\mathbf{x}) \leq 0. \end{aligned}$$

约束不等式 $g(\mathbf{x}) \leq 0$ 称为原始可行性(primal feasibility)，据此我们定义可行域(feasible region) $K = \mathbf{x} \in \mathbb{R}^n | g(\mathbf{x}) \leq 0$ 。假设 $\mathbf{x}^\star$ 为满足约束条件的最佳解，分开两种情况讨论：

(1) $g(\mathbf{x}^\star) < 0$ ，最佳解位于 K 的内部，称为内部解(interior solution)，这时约束条件是无效的(inactive)；

(2) $g(\mathbf{x}^\star) = 0$ ，最佳解落在 K 的边界，称为边界解(boundary solution)，此时约束条件是有效的(active)。

这两种情况的最佳解具有不同的必要条件。

(1)内部解：在约束条件无效的情形下， $g(\mathbf{x})$ 不起作用，约束优化问题退化为无约束优化问题，因此驻点 $\mathbf{x}^\star$ 满足 $\nabla f = \mathbf{0}$ 且 $\lambda = 0$ 。

(2)边界解：在约束条件有效的情形下，约束不等式变成等式 $g(\mathbf{x}) = 0$ ，这与前述Lagrange乘数法的情况相同。我们可以证明驻点 $\mathbf{x}^\star$ 发生于 $\nabla f \in \operatorname{span} \nabla g$ ，换句话说，存在 λ 使得 $\nabla f = -\lambda \nabla g$ ，但这里 λ 的正负号是有其意义的。**因为我们希望最小化 f ，梯度 ∇f (函数 f 在点 $\mathbf{x}$ 的最陡上升方向)应该指向可行域 K 的内部(因为你的最优解最小值是在边界取得的)，但 ∇g 指向 K 的外部(即 $g(\mathbf{x}) > 0$ 的区域，因为你的约束是小于等于0)，因此 $\lambda \geq 0$ ，称为对偶可行性(dual feasibility)。**

因此，不论是内部解或边界解， $\lambda g(\mathbf{x}) = 0$ 恒成立，称为互补松弛性(complementary slackness)。整合上述两种情况，最佳解的必要条件包括Lagrangian函数 $L(\mathbf{x}, \lambda)$ 的定常方程式、原始可行性、对偶可行性，以及互补松弛性：

优化模型 Find w 和 b

$$\begin{cases} \min: \Phi(w) = \frac{1}{2} w^T w \\ \text{Subject to}: d_i (w^T x_i + b) \geq 1 \end{cases}$$

对偶问题推导

$$\min \quad L(w, b, \lambda) = \frac{1}{2} w^T w - \sum_i \lambda_i [d_i (w^T x_i + b) - 1]$$

$$\text{s.t.} \quad \frac{\partial L(w, b, \lambda)}{\partial w} = 0 \Rightarrow w = \sum_i \lambda_i d_i x_i$$

$$\frac{\partial L(w, b, \lambda)}{\partial b} = 0 \Rightarrow \sum_i \lambda_i d_i = 0$$

$$\lambda_i \geq 0$$

$$d_i(w^T x_i + b) \geq 1$$

$$\lambda_i [d_i(w^T x_i + b)] = 0$$

$\Rightarrow$ $L(w, b, \lambda) = \sum_i \lambda_i - \frac{1}{2}\sum_i \sum_j \lambda_i d_i \lambda_j d_j x_j^T x_i$

s.t. $\begin{cases} \lambda_i \geq 0 \\ \sum_i \lambda_i d_i = 0 \end{cases}$

求解得到 λ_i, d_i 后

$$w = \sum_i \lambda_i d_i x_i$$

$b = 1 - w x_i$ （x_i 应为支撑向量且 $d_i = 1$）

$$d_i(w^T x_i + b) = 1$$

SVM推导

超平面 $w^T \vec{x} + b = 0$

$$g(x) = w^T x + b$$

$$g(x) = w^T(x - \vec{x})$$

$$= w^T r$$

$$g(x) = \|w\| r$$

$$r \propto \frac{1}{\|w\|}$$

$\Rightarrow$ $\min \frac{1}{2} w^T w$

subject to $d_i(w^T x_i + b) \geq 1$ for $i = 1:N$

引入拉格朗日乘子

$\min L(w, b) = \frac{1}{2} w^T w - \sum_{i=1}^{N} \lambda_i [d_i(w^T x_i + b) - 1]$

$$\begin{cases} \dfrac{\partial L(w.b)}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{N} \lambda_i d_i x_i \\ \dfrac{\partial L(w.b)}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N} \lambda_i d_i = 0 \\ \lambda_i \geq 0 \qquad d_i(w^T x_i + b) - 1 \geq 0 \end{cases}$$

$$\Rightarrow \quad = \frac{1}{2}\left(\sum_i \lambda_i d_i x_i^T\right)\left(\sum_j \lambda_j d_j x_j\right) + \sum_i \lambda_i - \sum_i \lambda_i d_i b - \sum_i \lambda_i d_i \left(\sum_j \lambda_j d_j x_j^T\right) x_i$$

$$\min \sum_i \lambda_i - \frac{1}{2} \sum_i \sum_j \lambda_i d_i \lambda_j d_j x_i^T x_j$$

$$\begin{cases} \lambda_i \geq 0 \\ \sum_i \lambda_i d_i = 0 \end{cases}$$

在解出 λ_i 之后

$$w = \sum_i \lambda_i d_i x_i$$

$$b = d_i - w^T x_i \quad (x_i \text{支撑向量})$$