训练数据集：模拟了问题的输入和输出
决策函数：告诉决策函数输入和输出
告诉函数：只给一函数
优化算法：计算优化问题最优解
超参数调参

REGRESSION TREE.

$\{(x_1, y_1)\}_{i=1}^{n}$ $\quad x_i = (x_{i1}, x_{i2}, \cdots x_{ij})$.

LOSS: $\sum_{i=1}^{n} (f(x_i) - y_i)^2$. $\quad R_1 = \{x \mid x_j < s\}$, $R_2 = \{x \mid x_j \geq s\}$

Opt: Greedy

Family: 线性 $f(x) = \sum_{i=1}^{n} c_i I_{\{x \in R_i\}}$

$$\min_{j,s,c_1,c_2} \sum_{i=1}^{n} (f(x_i) - y_i)^2 \Leftrightarrow \min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right]$$

$\hat{c}_1 = \text{mean}(y_i \mid x_i \in R_1(j,s))$

$\hat{c}_2 = \text{mean}(y_i \mid x_i \in R_2(j,s))$

s：某变量优化。

x_{ij} → 对 x_i 的 j 个坐标问题。

逻辑回归与回归 重点.

① 输出概率值 阈值
② 损失角度 上界
③ 思维区别.

CLASSIFICATION TREE.

R 划 $R_1, \cdots, R_m$，预测结果 $\hat{c}_1, \cdots, \hat{c}_m$

类别概率 $\hat{p}_{mk} = \frac{1}{N_m} \sum_{x_i \in R_m} I(y_i = k)$ $\quad$ m棵决策树第k个结果

结果 $k = \arg\max_k \hat{p}_{mk}$

LOSS 错误率: $\sum_{m=1}^{M} 1 - \hat{p}_{mk}$

基尼 系数: $\sum_{m=1}^{M} \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk})$.

交叉熵: $\sum_{m=1}^{M} \sum_{k=1}^{K} -\hat{p}_{mk} \log \hat{p}_{mk}$

聚类

① 聚类对象 → 空间聚类.
② 聚类的目的
③ 聚类的 距离
④ 聚类

相似度度量
L_p, Pearson, cos

层次聚类 凝聚 聚类 树.
① 合并法
k 个聚类 → k，每次从 近的两簇 开始合并
② 分裂法
k 个聚类 → k，选择最佳的划分方式 对每个 k.

算法

K-means.

K 初始化 μ_k → 迭代 → 计算 E_{ik}，K，均值.

μ_k

随机选择 K 个点

μ_k 的值 $\Leftarrow$ 非凸

高维度量.
$(x_1, \cdots, x_n) \to k$ 簇 $C = \{C_1, C_2, \cdots, C_k\}$.
$\mu = (\mu_1, \cdots, \mu_k)$.

$$\text{s.t.} \arg\min_{\mu, C} \sum_{i=1}^{k} \sum_{x_j \in C_i} \| x_j - \mu_i \|_2^2$$

表示数据集中一簇数据，最小
$\underbrace{\text{固定}\mu, \text{优化 } C}_{\text{固定}\mu, \text{优化 } C} \to \text{固定} C (\text{优化 } \mu)$ → 迭代收敛

$f(x) = \arg\min_i \| x - \mu_i \|^2$ 决策函数集.

GMM.

K个混合分布.
第 i 个分布的权重系数.
概率 $\pi_i = P(y = i)$. 选择第 i 个混合分布
概率 $N(\mu_i, \Sigma_i)$ 的高斯分布
生成模型. 易被 $\{y\}$ 每个数据的概率 y 分布
$\mathbb{P}\{x \mid \underbrace{y = i}_{\text{隐变量}}\} = N(\mu_i, \Sigma_i)$.

$$p(x) = \underbrace{\sum_{i=1}^{K}}_{\text{求和概率}} \underbrace{p(x \mid y = i)}_{\text{选择分布}} \underbrace{P(y = i)}_{\text{选择概率}}$$

隐变量 $\log \frac{P\{y=i \mid x\}}{P\{y=j \mid x\}} = \log \frac{P\{x \mid y=i\} P\{y=i\} / p(x)}{P\{x \mid y=j\} P\{y=j\} / p(x)}$

$= \log \frac{P\{x \mid y=i\} \pi_i}{P\{x \mid y=j\} \pi_j}$ $\quad (x, N(\mu, \Sigma))$

$= x^T W x + w^T x + C$. $\quad$ 二阶决策面

若 $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_K = \sigma^2 I$ 则 x 的决策面 线性 决策面

极大: $\arg\max_\theta \prod_{j=1}^{n} P\{x_j \mid \theta\} = \arg\max_\theta \prod_{j=1}^{n} \sum_{i=1}^{K} P\{x_j = i, x_j \mid \theta\}$

$$\boxed{\theta = [\mu_1, \cdots, \mu_k, \Sigma_1, \cdots, \Sigma_k, \pi_1, \cdots, \pi_k]} = \arg\max_\theta \prod_{i=1}^{n} \sum_{i=1}^{K} \mathbb{P}\{y_j = i \mid \theta\} P(x_j \mid y_j = i, \theta)$$

$$= \arg\max_\theta \prod_{j=1}^{n} \sum_{i=1}^{K} \pi_i \frac{1}{\sqrt{|2\pi|\Sigma|}} \exp\left\{ -\frac{1}{2} (x_j - \mu_i)^T \Sigma_i^{-1} (x_j - \mu_i) \right\}$$

EM算法:
$$\arg\max_\theta \prod_j P(x_j, y_j \mid \theta) \xrightarrow{MLE} \arg\max_\theta \sum_j \log P(x_j, y_j \mid \theta)$$

E-step 算. $\mathbb{E}_{Z \sim P(Z \mid X, \theta)} \log \mathbb{P}\{X, Y \mid \theta\}$

$$= \sum_{j=1}^{K} \underbrace{P(Y_j = j \mid x_i)}_{\text{选择概率}} \sum_{i=1}^{n} \underbrace{\log P(X_i, Y_i = j \mid \theta)}_{\text{选择概率的联合概率计算}} = Q(\theta^t \mid \theta^{t-1})$$

M-step，求偏导 最大化 $Q(\theta^t \mid \theta^{t-1})$.

对任一分布 $g(x)$. $\quad$ 求一.

$$\log P(D \mid \theta^t) = \int dy\, g(y) \log P(D \mid \theta^t)$$

$$= \int dy\, g(y) \log \frac{P(y, D \mid \theta^t)\, g(y)}{P(y \mid D, \theta^t)\, g(y)}$$

$$= \underbrace{\int dy\, g(y) \log P(y, D \mid \theta^t) - \int g(y)\, dy \log g(y)}_{\mathcal{F}_{\theta^t}(g(\cdot), D)} + \underbrace{\int dy\, g(y) \log \frac{g(y)}{P(y \mid D, \theta^t)}}_{D_{KL}(g(y) \| P(y \mid D, \theta^t))}$$

$\geq \mathcal{F}_{\theta^t}(g(\cdot), D)$ $\quad$ 设 $g(\cdot)$ 最大化 $\mathcal{F}_{\theta^t}$ $\Leftrightarrow$ 最小化 $D_{KL}(g(y) \| P(y \mid D, \theta^t))$.

令 $g(\cdot) = p(y \mid D, \theta^t)$.

线性最小二乘.

$$\theta_{MLE} = \arg\max_\theta \mathbb{P}\{\text{one label from one}\} = \arg\max_\theta \mathbb{P}\{D \mid \theta\}$$

Family: $\mathbb{P}\{y \mid x; w\}$ $\quad (\sim N(y \mid f_w(x), \sigma^2), y = f_w(x) + \varepsilon)$.

LOSS: $\max_w p(Y \mid X; w)$ $\quad (\max \prod_i N(y_i \mid f_w(x_i), \sigma^2))$

Opt: Least Square

BAYES

$$P(\theta \mid D\} = \frac{P\{D \mid \theta\} P\{\theta\}}{P\{D\}}$$

$\xrightarrow{\text{Bias}}$ $\quad$ 高斯分布 $\{\#j : Y^{(j)} = y\}/n$.

$\sim P\{D \mid \theta\} P\{\theta\}$

likelihood $\xleftarrow{P(x \mid \theta)}$ $\frac{\{\#j : X^{(j)} = x_i, Y^{(j)} = y\}/n}{\{\#j : Y^{(j)} = y\}/n}$

MAP:

Family: $P\{y \mid x; w\}$ $p(w)$

LOSS: $\max_w p(Y \mid X; w) p(w)$.

Opt: 极大 最小

决策

分类. ① 生命 $\quad$ 天气.
② 医生 $\quad$ 若癌，生命危险 需要治疗.
③ 医生 $\quad$ 治疗 $\quad$ 生命.

R中的 ① 回归.
最终回归函数 (LS.T). $\quad$ 多为 $p(x \mid Y) = p(Y \mid x) p(x)$

$E_{p(x,y)}(Y) \to$ Bias + Var 分解.

R中的 分类.
比较 $\mathbb{P}(X = x \mid Y = y_1) P(Y = y_1)$

$\min_f \int_x \int_y (f(x) - y_i)^2 p(y \mid x) p(x) dx dy$.

$f^* = \arg\min_f \underbrace{P(f(x) \neq Y)}_{\text{期望风险最小}} \Leftrightarrow f^*(x) = \arg\max_{Y \in y} P(Y = y \mid X = x)$.

$\overset{\text{Bayes}}{=} \arg\max_{Y \in y} P\{X = x \mid Y = y\} P\{Y = y\}$. $\quad$ 条件概率 最大 $\quad$ 最大概率

条件独立. $\forall x, y, z$. $P(X = x \mid Y = y, Z = z) = P(X = x \mid Z = z)$.

Family: $p(Y \mid X, \theta) \xrightarrow{\text{条件概率}} \prod_{i=1}^{n} p(y^i \mid x^i, \theta_i)$.

LOSS: $\mathbb{P}\{Y = x \mid Y = y\} \xrightarrow{\text{Naive}} \prod_{i=1}^{n} \mathbb{P}\{X_i = x_i \mid Y = y\}$. $\quad$ Bayes. $\quad$ 样本出现概率 概率

K 类 K-计算.
总共 $(2^d - 1)K \to$ 减少

MLE → MAP.
假设存在 个 θ，使 X 出现概率最大.
增加 θ 先验 概率

PCA.

最大 距方差.
数据 $\{x_i\}$ 处理.
第一主成分: 最大方差的数据 轴.
第一主成分: 数据在第一主成分上 投影 多
之后
① 特征 与 特征 的 计算.
② 数据 投影 方差 尽可能 最大.

$$w_1 = \arg\max_{\|w\|=1} \frac{1}{m} \sum_{i=1}^{m} (w^T x_i)^2$$

$$\Leftrightarrow \begin{cases} \max w^T X X^T w \\ \text{s.t. } w^T w = 1 \end{cases}$$

$L(w, \lambda) = w^T \Sigma w - \lambda(w^T w - 1)$.

$\frac{\partial L}{\partial w} = \frac{1}{2} \Sigma w - \frac{1}{2} \lambda w = 0$.

$\Sigma w = \lambda w$. $\quad$ λ 为 Σ 特征值. w 为特征向量

$w^T(\lambda w + 2\lambda w_1)$
$= \lambda$.

$w^T w = 1$. $\quad$ s.t. $w^T w = 1$.
$w_1 w = 0$. $\quad w_1, \cdots, w_{k-1}$.

高维数据 降维 在 空间 几何 性质 是 反直

要 降空间 改 一个 原 空间

ISOMAP: 邻域图 → 新 点对 计算 最短 路径 → MDS 降 维

local linear embedding. 保持 局部 线性
$\min \| x_i - \sum_{j \in N_i} w_{ij} x_j \|^2$. 若 x_1, x_2 多 邻域 则 $w_{ij} \approx 0$
s.t. $\sum_j w_{ij} = 1$.

高维数据 降维 在 分类

一个 低维 空间 不 分 $\to$ 升维 $\to$ 线性 可分.
E 中 - 给 量 核 增 …

Auto encoder: AKPCA → AE.
线性 → 非线性
单隐层 → 多隐层

VAE 生成 优化 KL 散度

MDS

$D \in \mathbb{R}^{n \times n}$. dist_{ij} 为 欧氏距离
求 $Z \in \mathbb{R}^{d' \times m}$. $d' < d$ 且使 两个 距离 在 低维空间 和 原空间 欧氏距离 尽可能 相等

$\|z_i - z_j\| = \text{dist}_{ij}$. $\quad \sum_i z_i = 0$. $\quad \sum_i z_i = 0$.

令 $B = Z^T Z \in \mathbb{R}^{m \times m}$. $\quad b_{ij} = z_i^T z_j$

$\text{dist}_{ij}^2 = \|z_i\|^2 + \|z_j\|^2 - 2 z_i^T z_j$
$= b_{ii} + b_{jj} - 2b_{ij}$

$\sum_{i=1}^{m} b_{ij} = \sum_{j=1}^{m} b_{ij} = 0$.

令 $\sum_{i=1}^{m} \text{dist}_{ij}^2 = tr(B) + m b_{jj}$.
$\sum_{j=1}^{m} \text{dist}_{ij}^2 = tr(B) + m b_{ii}$
$\sum_{i=1}^{m} \sum_{j=1}^{m} \text{dist}_{ij}^2 = 2m\, tr(B)$

令 $tr(B) = \sum_{i=1}^{m} \|z_i\|^2$

$\text{dist}_{i\cdot}^2 = \frac{1}{m} \sum_{j=1}^{m} \text{dist}_{ij}^2$

$\text{dist}_{j\cdot}^2 = \frac{1}{m} \sum_{i=1}^{m} \text{dist}_{ij}^2$

$\text{dist}^2 = \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \text{dist}_{ij}^2$.

故 $b_{ij} = -\frac{1}{2}(\text{dist}_{ij}^2 - \text{dist}_{i\cdot}^2 - \text{dist}_{j\cdot}^2 + \text{dist}^2)$

故 D 计算 B.
令 $B = V \Lambda V^T$. 取 d' 个 特征值 与 特征 向量 的 乘积

$Z = \tilde{\Lambda}^{1/2} \tilde{V}^T \in \mathbb{R}^{d' \times m}$.

感知机.

$\tilde{w}^T \tilde{x} + b = 0$

$\tilde{w}$ 法向量. $\quad b$ 位移 (决定距离).

超平面 中任 点. $\tilde{x}$. 到 超平面 $(\tilde{w}, b)$ 的距离.

$$r = \frac{|\tilde{w}^T \tilde{x} + b|}{\|\tilde{w}\|}$$

类 $(\tilde{w}, b)$ 感知机 满足 正确 决策

即 对. $(\tilde{x}_i, y_i) \in D$. $\quad y_i = +1 \Rightarrow \tilde{w}^T \tilde{x}_i + b > 0$.
$y_i = -1 \Rightarrow \tilde{w}^T \tilde{x}_i + b < 0$.

令 $\begin{cases} \tilde{w}^T \tilde{x}_i + b \geq 1 & y_i = +1 \\ \tilde{w}^T \tilde{x}_i + b \leq -1 & y_i = -1 \end{cases}$

支持向量 满足等号 的点 是 最近点.

$\hat{\gamma} = \frac{2}{\|\tilde{w}\|}$ 最 为 间隔. 支持向量 到 超平面 的 距离

支持 最大间隔 超平面

Find w, b. s.t. $\hat{\gamma}$ 最大

$\max_{\tilde{w}, b} \frac{2}{\|\tilde{w}\|}$

s.t. $y_i(\tilde{w}^T \tilde{x}_i + b) \geq 1$. $\quad i = 1, 2, \cdots, m$

$\Updownarrow$

$\min_{\tilde{w}, b} \frac{1}{2} \|\tilde{w}\|^2$

s.t. $y_i(\tilde{w}^T \tilde{x}_i + b) \geq 1$. $\quad i = 1, 2, \cdots, m$

对偶

$L(w, b, \alpha) = \frac{1}{2} \|\tilde{w}\|^2 - \sum_{i=1}^{m} \alpha_i (y_i(\tilde{w}^T \tilde{x}_i + b) - 1)$

令偏导为 0.

$\frac{\partial L}{\partial \tilde{w}} = 0$. $\quad \frac{\partial L}{\partial b} = 0$.

$\Downarrow$

$\tilde{w} = \sum_{i=1}^{m} \alpha_i y_i \tilde{x}_i$

$0 = \sum_{i=1}^{m} \alpha_i y_i$

代回 式 $\max_\alpha L(\tilde{w}, b, \tilde{x})$

s.t. $\sum_{i=1}^{m} \alpha_i y_i = 0$.

$\alpha_i \geq 0$. $\quad i = 1, \cdots, m$

解 y_i α_i 乘 $\tilde{x}$.

KKT.

$$\begin{cases} \alpha_i \geq 0 \\ y_i f(\tilde{x}_i) - 1 \geq 0 \\ \alpha_i (y_i f(\tilde{x}_i) - 1) = 0 \end{cases}$$

即 若 $\alpha_i = 0$. 则 对 $f(\tilde{x})$ 无贡献
$\alpha_i > 0$. 则 必有 $y_i f(\tilde{x}) = 1$.
此时 得 与 最 近 点 在 最大 边界 上.
只 支持 向 量.

$R(f) < R_{emp}(f) + \phi(n, VC(H))$

大数. $n \to \infty$.
$\phi \to 0$.
ERM → 最小

情况. 经验风险 小. 泛化 风险 大.

SRM. $\min_H VC(H)$ $\quad$ 经验风险 小 和 置信 风险 $\quad$ emp L.
s.t. $R_{emp}(f)$ 最小. $\quad$ 经验风险 $\quad$ VC(H) L

VC 维.

H 集合 在 点 集 s.t. $(x_1, \cdots, x_n)$ 的 分类
可能 被 分 成 真 到 2^n 种 打散
若 H 能打散 任意 数据 集 的 分类 则 VC 维.

若 线性 不 可分. ① 软化.

模型 $f(\tilde{x}) = \tilde{w}^T \phi(\tilde{x}) + b$

$\min_{\tilde{w}, b} \frac{1}{2} \|\tilde{w}\|^2$

s.t. $y_i(\tilde{w}^T \phi(\tilde{x}_i) + b) \geq 1$. $\quad i = 1, 2, \cdots, m$.

其对偶问题 为

$$\max_\alpha \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \phi(\tilde{x}_i)^T \phi(\tilde{x}_j)$$

s.t. $\sum_{i=1}^{m} \alpha_i y_i = 0$

$\alpha_i \geq 0$. $\quad i = 1, 2, \cdots, m$

$\phi(\tilde{x}_i)^T \phi(\tilde{x}_j)$. 计算 困难

$K(\tilde{x}_i, \tilde{x}_j) = \langle \phi(\tilde{x}_i), \phi(\tilde{x}_j) \rangle$

解出 $\tilde{w}$.

$f(\tilde{x}) = \sum_{i=1}^{m} \alpha_i y_i K(\tilde{x}, \tilde{x}_i) + b$

核函数.

$\mathcal{X}$ 输入空间. $\quad K(\cdot, \cdot)$ $\quad \mathcal{X} \times \mathcal{X}$ 对称.

K 是 核函数 当 会 当 时 任意 数据. $D = \{\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_m\}$.

核 矩阵 是 半正定 的

$$K = \begin{bmatrix} K(\tilde{x}_1, \tilde{x}_1) & \cdots & K(\tilde{x}_1, \tilde{x}_j) & \cdots & K(\tilde{x}_1, \tilde{x}_m) \\ \vdots & \ddots & & & \vdots \\ K(\tilde{x}_i, \tilde{x}_1) & \cdots & K(\tilde{x}_i, \tilde{x}_j) & \cdots & K(\tilde{x}_i, \tilde{x}_m) \\ \vdots & & & \ddots & \vdots \\ K(\tilde{x}_m, \tilde{x}_1) & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

对任意一个半正定 核 矩阵 总能找到 映射 唯一的 ϕ.

$\phi \to$ 再生 核希尔 空间 (RKHS).

核 函数 表示

③ 软间隔 + 正则.

$$\min_{\tilde{w}, b, \xi} \frac{1}{2} \|\tilde{w}\|^2 + C \sum_{i=1}^{m} \ell_{0/1}(y_i(\tilde{w}^T \tilde{x}_i + b) - 1)$$

$C > 0$ 常数. $\quad \ell_{0/1}(z) = \begin{cases} 1, & z < 0 \\ 0 & \text{otherwise} \end{cases}$

Hinge 损失 替代 $\ell_{0/1}(z)$.

$\min_{\tilde{w}, b} \frac{1}{2} \|\tilde{w}\|^2 + C \sum_{i=1}^{m} \max(0, 1 - y_i(\tilde{w}^T \tilde{x}_i + b))$.

$\Downarrow$ 引入 松弛 变量. $\xi_i \geq 0$.
ξ_i 为 每个 样本 点 的 损失 函数

$\min_{\tilde{w}, b, \xi} \frac{1}{2} \|\tilde{w}\|^2 + C \sum_{i=1}^{m} \xi_i$.

s.t. $y_i(\tilde{w}^T \tilde{x}_i + b) \geq 1 - \xi_i$

$\xi_i \geq 0$. $\quad i = 1, 2, \cdots, m$.

结构风险. $\quad$ 经验风险.

$\underbrace{\Omega(f)}_{\text{正则项}} + C \underbrace{\sum_{i=1}^{m} \ell(f(x_i), y_i)}_{\text{损失函数}}$.

BOOSTING.

= 弱. 基模型 ε. $\forall h_i$.
$\{x, y\}$.

$H(\tilde{x}) = \text{sign}\left(\sum_{i=1}^{T} h_i(\tilde{x})\right)$

独立. 每个 变量 步骤 independent

$\mathbb{P}\{H(\tilde{x}) \neq f(\tilde{x})\} = \sum_{k=\frac{T}{2}}^{T} \binom{T}{k} (1-\varepsilon)^k \varepsilon^{T-k}$.

$\leq \exp\left(-\frac{1}{2} T (1 - 2\varepsilon)^2\right)$.

Adaboost.

集 学习 的 自适应 增强. $H(\tilde{x}) = \sum_{i=1}^{T} \alpha_i h_i(\tilde{x})$

最 小化 指数 函数 $\quad \text{loss}(H \mid D) = \mathbb{E}_{\tilde{x} \sim D}[e^{-f(\tilde{x}) H(\tilde{x})}]$

$D = \{(\tilde{x}_1, y_1), \cdots, (\tilde{x}_m, y_m)\}$.

$\mathcal{L}$, T

$D_1(\tilde{x}) = \frac{1}{m}$.

For $t = 1, 2, \cdots, T$ do

$h_t = \mathcal{L}(D, D_t)$.

$\varepsilon_t = \mathbb{P}_{\tilde{x} \sim D_t}(h_t(\tilde{x}) \neq f(\tilde{x}))$.

If $\varepsilon_t > 0.5$ break.

$\alpha_t = \frac{1}{2} \ln\left(\frac{1 - \varepsilon_t}{\varepsilon_t}\right)$

$$D_{t+1}(\tilde{x}) = \frac{D_t(\tilde{x})}{Z_t} \times \begin{cases} \exp\{-\alpha_t\}, & \text{if } h_t(\tilde{x}) = f(\tilde{x}) \\ \exp\{\alpha_t\}, & \text{if } h_t(\tilde{x}) \neq f(\tilde{x}) \end{cases}$$

$= \frac{D_t(\tilde{x}) \exp\{-\alpha_t f(\tilde{x}) h_t(\tilde{x})\}}{Z_t}$. $\quad Z_t = \sum_{i=1}^{m} D_t(\tilde{x}) \exp\{-\alpha_t y_i h_t(\tilde{x})\}$.

End for.

$\Rightarrow H(\tilde{x}) = \text{sign}\left(\sum_{k=1}^{T} \alpha_k h_k(\tilde{x})\right)$ $\quad$ // 基模型. 令 寻找 α_t 最小化 Z_t.

推导 ①

$\text{loss}(\alpha_t h_t \mid D_t) = \mathbb{E}_{\tilde{x} \sim D_t}[e^{-f(\tilde{x}) \alpha_t h_t(\tilde{x})}]$

$= \mathbb{E}_{\tilde{x} \sim D_t}[e^{\alpha_t} \mathbb{1}\{f(\tilde{x}) = h_t(\tilde{x})\} + e^{-\alpha_t} \mathbb{1}\{f(\tilde{x}) \neq h_t(\tilde{x})\}]$

$= e^{-\alpha_t}(1 - \varepsilon_t) + e^{\alpha_t} \varepsilon_t$.

$\varepsilon_t = \mathbb{P}_{\tilde{x} \sim D_t}(h_t(\tilde{x}) \neq f(\tilde{x}))$. $\quad$ 对数 指数 求 偏导.

若对 α_t 偏导. 为

$-e^{-\alpha_t}(1 - \varepsilon_t) + e^{\alpha_t} \varepsilon_t$. $\quad$ 令正 收敛.

为零 得 $\alpha_t = \frac{1}{2} \ln\left(\frac{1 - \varepsilon_t}{\varepsilon_t}\right)$. $\Longrightarrow$ $Z_t = 2\sqrt{\varepsilon_t(1 - \varepsilon_t)}$

② 令 $H_{t-1}(\tilde{x})$. 为 前 t H_{t-1}.
h_t 最小化.

$\text{loss}(H_{t-1} + h_t \mid D) = \mathbb{E}_{\tilde{x} \sim D}[e^{-f(\tilde{x}) H_{t-1}(\tilde{x})} e^{-f(\tilde{x}) h_t(\tilde{x})}]$.

$\overset{\text{Taylor}}{\approx} \mathbb{E}_{\tilde{x} \sim D}\left[e^{-f(\tilde{x}) H_{t-1}(\tilde{x})} \left(1 - f(\tilde{x}) h_t(\tilde{x}) + \frac{f^2(\tilde{x}) h_t^2(\tilde{x})}{2}\right)\right]$

$= \mathbb{E}_{\tilde{x} \sim D}\left[e^{-f(\tilde{x}) H_{t-1}(\tilde{x})} \left(1 + \frac{1}{2} - f(\tilde{x}) h_t(\tilde{x})\right)\right]$

即. 最 大化.

$\mathbb{E}_{\tilde{x} \sim D}\left[\frac{e^{-f(\tilde{x}) H_{t-1}(\tilde{x})}}{\underbrace{\mathbb{E}_{\tilde{x} \sim D}[e^{-f(\tilde{x}) H_{t-1}(\tilde{x})}]}_{\text{与 } h_t \text{ 无关 } \to \text{ 常数}}} f(\tilde{x}) h_t(\tilde{x})\right]$

令 D_t 为一 分布. $\quad \frac{e^{-f(\tilde{x}) H_{t-1}(\tilde{x})}}{\mathbb{E}_{\tilde{x} \sim D}[e^{-f(\tilde{x}) H_{t-1}(\tilde{x})}]} \cdot D(\tilde{x})$.

则 $h_t = \arg\max_{h_t} \mathbb{E}_{\tilde{x} \sim D_t}[f(\tilde{x}) h(\tilde{x})]$ $\quad f(\tilde{x}) h(\tilde{x}) = 1 - 2\mathbb{1}\{f(\tilde{x}) \neq h(\tilde{x})\}$

$$D_{t+1}(\tilde{x}) = \frac{\overbrace{D(\tilde{x}) \frac{e^{-f(\tilde{x}) H_{t-1}(\tilde{x})}}{\mathbb{E}_{\tilde{x} \sim D}[e^{-f(\tilde{x}) H_{t-1}(\tilde{x})}]}}^{D_t(\tilde{x})} e^{-f(\tilde{x}) \alpha_t h_t(\tilde{x})}}{\underbrace{\mathbb{E}_{\tilde{x} \sim D}[e^{-f(\tilde{x}) H_{t-1}(\tilde{x})}]}_{\text{与 } h_t \text{ 无关}}}$$

$= D_t(\tilde{x}) \cdot \frac{e^{-f(\tilde{x}) \alpha_t h_t(\tilde{x})}}{\mathbb{E}_{\tilde{x} \sim D}[e^{-f(\tilde{x}) H_{t-1}(\tilde{x})}]}$

$f(x) = \text{sign}(\sum H_t(x))$

转换 损失 $\frac{1}{m} \sum_{i=1}^{m} \mathbb{1}\{\hat{f}(x) \neq y_i\}$
$\leq \frac{1}{m} \sum_{i=1}^{m} \exp\{-y_i H(x)\}$.
$= \prod_{t=1}^{T} Z_t$.
$= \prod_{t=1}^{T} 2\sqrt{\varepsilon_t(1 - \varepsilon_t)} = \sqrt{1 - (1 - 2\varepsilon_t)^2} < 1$.
$\leq \exp\left(-2 \sum_{t=1}^{T} (1/2 - \varepsilon_t)^2\right)$