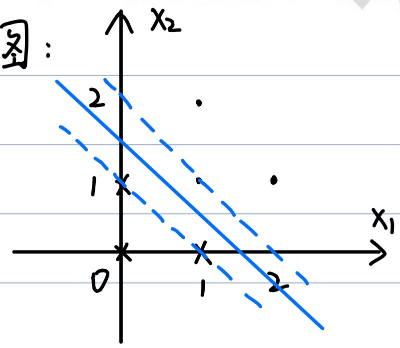
2. 一个样本集是

$$\boldsymbol{D}=\left\{\left((0,0)^{\mathrm{T}},-1\right),\left((1,0)^{\mathrm{T}},-1\right),\left((0,1)^{\mathrm{T}},-1\right),\left((1,1)^{\mathrm{T}},1\right),\left((1,2)^{\mathrm{T}},1\right),\left((2,1)^{\mathrm{T}},1\right),\left((2,2)^{\mathrm{T}},1\right)\right\}$$

训练一个 SVM 对其进行分类，写出判决方程，指出那些样本是支持向量。

作图：




判决方程：$x_1+x_2-\frac{3}{2}=0$

支持向量：$((1,0)^T,-1)$

$((0,1)^T,-1)$

$((1,1)^T,1)$

3. 异或的样本集 $\boldsymbol{D}=\left\{\left((0,0)^{\mathrm{T}},-1\right),\left((0,1)^{\mathrm{T}},1\right),\left((1,0)^{\mathrm{T}},1\right),\left((1,1)^{\mathrm{T}},-1\right)\right\}$ 是线性不可分的，

可定义一个多项式函数 $\boldsymbol{\varphi}(\boldsymbol{x})$，其中

$$\varphi_1(\boldsymbol{x})=2(x_1-0.5)$$
$$\varphi_2(\boldsymbol{x})=4(x_1-0.5)(x_2-0.5)$$

把样本 $\boldsymbol{D}=\left\{(\boldsymbol{x}_n,y_n)\right\}$ 映射成 $\boldsymbol{D}_{\varphi}=\left\{(\boldsymbol{\varphi}(\boldsymbol{x}_n),y_n)\right\}$，样本集映射为

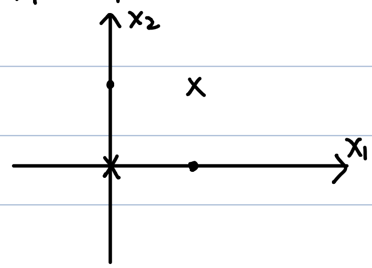
$$\boldsymbol{D}_{\varphi}=\left\{\left((-1,1)^{\mathrm{T}},-1\right),\left((-1,-1)^{\mathrm{T}},1\right),\left((1,-1)^{\mathrm{T}},1\right),\left((1,1)^{\mathrm{T}},-1\right)\right\}$$

分别用 $\boldsymbol{D}$ 样本集和 $\boldsymbol{D}_{\varphi}$ 样本集 $\boldsymbol{D}_{\varphi}$ 训练一个线性 SVM 分类器，并分析其分类性能（注：由于样本数少且取值简单，可手动练习，也可自行编写一段小程序实现，但不要使用机器学习的专用软件包）。

| $n$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $x_1$ | 0 | 0 | 1 | 1 |
| $x_2$ | 0 | 1 | 0 | 1 |
| $\varphi_1(x)=2(x_1-0.5)$ | -1 | -1 | 1 | 1 |
| $\varphi_2(x)=4(x_1-0.5)(x_2-0.5)$ | 1 | -1 | -1 | 1 |
| $y_n$ | -1 | 1 | 1 | -1 |

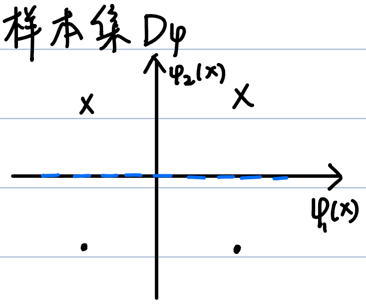
∴样本集 $D_\varphi$ 为 $\{(-1,1)^T,-1\},\{(-1,-1)^T,1\},\{(1,-1)^T,1\},\{(1,1)^T,-1\}$

样本集D：




线性不可分，得到的分类器性能差。

样本集$D_\varphi$




判决方程：$\varphi_2(x)=0$

$$4(x_1-0.5)(x_2-0.5)=0$$

线性可分，得到的分类器性能好。

4. 在引进松弛变量的 SVM 目标函数中，一种更一般的形式为

$$\min_{\boldsymbol{w},\xi_n}\left\{\frac{1}{2}\|\boldsymbol{w}\|^2\right\}+C\sum_{n=1}^{N}\xi_n^p$$

$$s.t.\begin{cases}\xi_n\geq 0\\ y_n\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_n+b\right)\geq 1-\xi_n\end{cases},\quad n=1,2,\cdots,N$$

只要 $p\geq 1$，则问题都是凸优化问题。本题取 $p=2$，这对应“平方合页损失”，在该情况下，推导其对偶优化表达式。

$$J=\min_{w,\xi_n}\left\{\frac{1}{2}\|w\|^2\right\}+C\sum_{n=1}^{N}\xi_n^2 \qquad s.t.\begin{cases}\xi_n\geq 0\\ y_n(w^Tx_n+b)\geq 1-\xi_n\end{cases},\ n=1,2,\cdots,N$$

拉格朗日函数

$$L(w,b,\xi_n,a,\mu)=\frac{1}{2}\|w\|^2+C\sum_{n=1}^{N}\xi_n^2-\sum_{n=1}^{N}a_n[y_n(w^Tx_n+b)-1+\xi_n]-\sum_{n=1}^{N}\mu_n\xi_n \quad ①$$

KKT条件：$a_n[y_n(w^Tx_n+b)-1+\xi_n]=0$

$\mu_n\xi_n=0$，$n=1,2,\cdots,N$

$$\frac{\partial L}{\partial w} = w - \sum_{n=1}^{N} a_n y_n x_n = 0 \qquad \Rightarrow w = \sum_{n=1}^{N} a_n y_n x_n \quad ②$$

$$\frac{\partial L}{\partial b} = -\sum_{n=1}^{N} a_n y_n = 0 \qquad \Rightarrow \sum_{n=1}^{N} a_n y_n = 0 \quad ③$$

$$\frac{\partial L}{\partial \xi_n} = 2C\xi_n - a_n - \mu_n = 0 \qquad \Rightarrow a_n = 2C\xi_n - \mu_n \quad ④$$

②③④代入①：

$$L(a) = \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m y_n y_m \langle x_n, x_m \rangle + \sum_{n=1}^{N} a_n - 2C\xi_n^2$$

$$\text{s.t.} \sum_{n=1}^{N} a_n y_n = 0$$

1. 样本集 $\boldsymbol{D}$ 共有 20 个样本，类型只有正样和负样两类，其中正样 9 个，负样 11 个。有一个特征 $A$，其取值分别为 $i=1,2,3$，按照该特征的不同取值将样本集 $\boldsymbol{D}$ 划分为子集 $\boldsymbol{D}_i, i=1,2,3$，各子集的样本数分别为 $N_1=8, N_2=7, N_3=5$，并且 $\boldsymbol{D}_1$ 有正样 2 个，$\boldsymbol{D}_2$ 有正样 3 个
   (1) 求通过特征 $A$ 得到的信息增益和信息增益率。
   (2) 求 $A=2$ 的基尼指数。

(1) $H(D) = -\frac{9}{20}\log\frac{9}{20} - \frac{11}{20}\log\frac{11}{20} = 0.6881$

$$H(D|A) = \sum P_i^{(A)} H(D|A=i)$$

$$= \frac{8}{20}\left(-\frac{2}{8}\log\frac{2}{8} - \frac{6}{8}\log\frac{6}{8}\right) + \frac{7}{20}\left(-\frac{3}{7}\log\frac{3}{7} - \frac{4}{7}\log\frac{4}{7}\right)$$

$$+ \frac{5}{20}\left(-\frac{4}{5}\log\frac{4}{5} - \frac{1}{5}\log\frac{1}{5}\right)$$

$$= 0.5890$$

$$G(D,A) = H(D) - H(D|A) = 0.0991$$

(2) $Gini(D) = 1 - \sum_{k=1}^{K} P_k^2$

$$= \frac{7}{20}\left(1 - \left(\frac{3}{7}\right)^2 - \left(\frac{4}{7}\right)^2\right) + \frac{13}{20}\left(1 - \left(\frac{6}{13}\right)^2 - \left(\frac{7}{13}\right)^2\right)$$

$$= 0.4945$$

4. 如下是贷款样本集，分别使用 ID3 算法和 CART 分类树算法给出判断是否贷款得如决策

| ID | AGE | JOB_STATUS | OWNS_HOUSE | CREDIT_RATING | CLASS (Yes or No) |
|---|---|---|---|---|---|
| 1 | Young | ~~False~~ | False | Fair | No |
| 2 | Young | False | False | Good | No |
| 3 | Young | ~~True~~ | False | Good | Yes |
| 4 | Young | ~~True~~ | True | ~~Fair~~ | ~~Yes~~ |
| 5 | Young | ~~False~~ | False | Fair | No |
| 6 | Middle | ~~False~~ | False | Fair | No |
| 7 | Middle | False | False | Good | No |
| 8 | Middle | ~~True~~ | True | ~~Good~~ | ~~Yes~~ |
| 9 | Middle | False | True | ~~Excellent~~ | ~~Yes~~ |
| 10 | Middle | False | True | ~~Excellent~~ | ~~Yes~~ |
| 11 | Old | False | True | ~~Excellent~~ | ~~Yes~~ |
| 12 | Old | False | True | ~~Good~~ | ~~Yes~~ |
| 13 | Old | ~~True~~ | False | Good | Yes |
| 14 | Old | ~~True~~ | False | ~~Excellent~~ | ~~Yes~~ |
| 15 | Old | ~~False~~ | False | Fair | No |
| | A | B | C | E | 9, 6 |

ID3:

$$H(D) = -\frac{9}{15}\log\frac{9}{15} - \frac{6}{15}\log\frac{6}{15} = 0.6730$$

$$H(D|A) = \frac{5}{15}\left(-\frac{2}{5}\log\frac{2}{5} - \frac{3}{5}\log\frac{3}{5}\right) + \frac{5}{15}\left(-\frac{2}{5}\log\frac{2}{5} - \frac{3}{5}\log\frac{3}{5}\right) + \frac{5}{15}\left(-\frac{4}{5}\log\frac{4}{5} - \frac{1}{5}\log\frac{1}{5}\right)$$

$$= 0.6155$$

$$H(D|B) = \frac{5}{15}\left(-\frac{5}{5}\log\frac{5}{5} - \frac{0}{5}\log\frac{0}{5}\right) + \frac{10}{15}\left(-\frac{4}{10}\log\frac{4}{10} - \frac{6}{10}\log\frac{6}{10}\right)$$

$$= 0.4487$$

$$H(D|C) = \frac{6}{15}\left(-\frac{6}{6}\log\frac{6}{6} - \frac{0}{6}\log\frac{0}{6}\right) + \frac{9}{15}\left(-\frac{3}{9}\log\frac{3}{9} - \frac{6}{9}\log\frac{6}{9}\right)$$

$$= 0.3819$$

$$H(D|E) = \frac{5}{15}\left(-\frac{1}{5}\log\frac{1}{5} - \frac{4}{5}\log\frac{4}{5}\right) + \frac{6}{15}\left(-\frac{4}{6}\log\frac{4}{6} - \frac{2}{6}\log\frac{2}{6}\right) + \frac{4}{15}\left(-\frac{4}{4}\log\frac{4}{4} - \frac{0}{4}\log\frac{0}{4}\right)$$

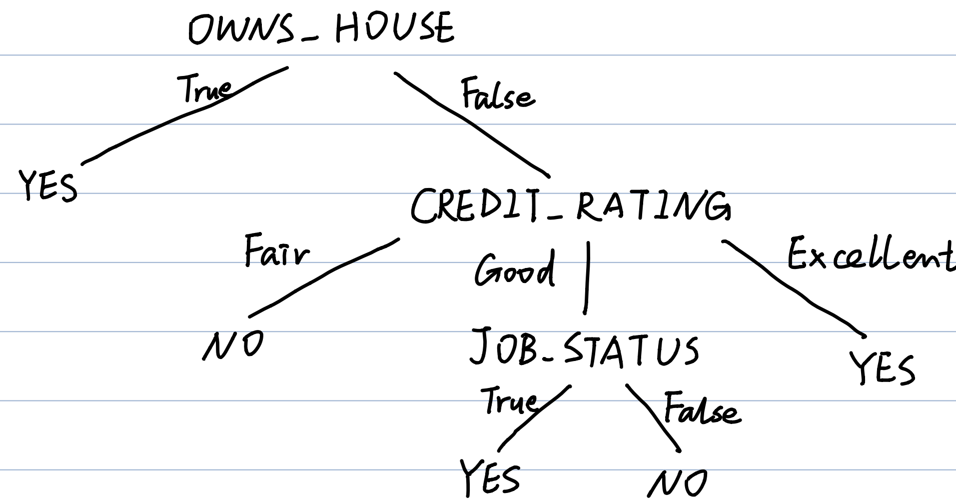
$$= 0.4214$$

$$G(D,A) = H(D) - H(D|A) = 0.0275$$

$$G(D,B) = H(D) - H(D|B) = 0.2243$$

$$G(D,C) = H(D) - H(D|C) = 0.2911$$

$$G(D,E) = H(D - H(D|E) = 0.2516$$




CART:

$$Gini(D, A=1) = \frac{5}{15}\left(1-\left(\frac{2}{5}\right)^2-\left(\frac{3}{5}\right)^2\right) + \frac{10}{15}\left(1-\left(\frac{7}{10}\right)^2-\left(\frac{3}{10}\right)^2\right)$$
$$= 0.44$$

$$Gini(D, A=2) = \frac{5}{15}\left(1-\left(\frac{3}{5}\right)^2-\left(\frac{2}{5}\right)^2\right) + \frac{10}{15}\left(1-\left(\frac{6}{10}\right)^2-\left(\frac{4}{10}\right)^2\right)$$
$$= 0.48$$

$$Gini(D, A=3) = \frac{5}{15}\left(1-\left(\frac{4}{5}\right)^2-\left(\frac{1}{5}\right)^2\right) + \frac{10}{15}\left(1-\left(\frac{5}{10}\right)^2-\left(\frac{5}{10}\right)^2\right)$$
$$= 0.44$$

$\therefore A = 1$

$$Gini(D, B) = \frac{5}{15}(1-1-0) + \frac{10}{15}\left(1-\left(\frac{4}{10}\right)^2-\left(\frac{6}{10}\right)^2\right) = 0.32$$

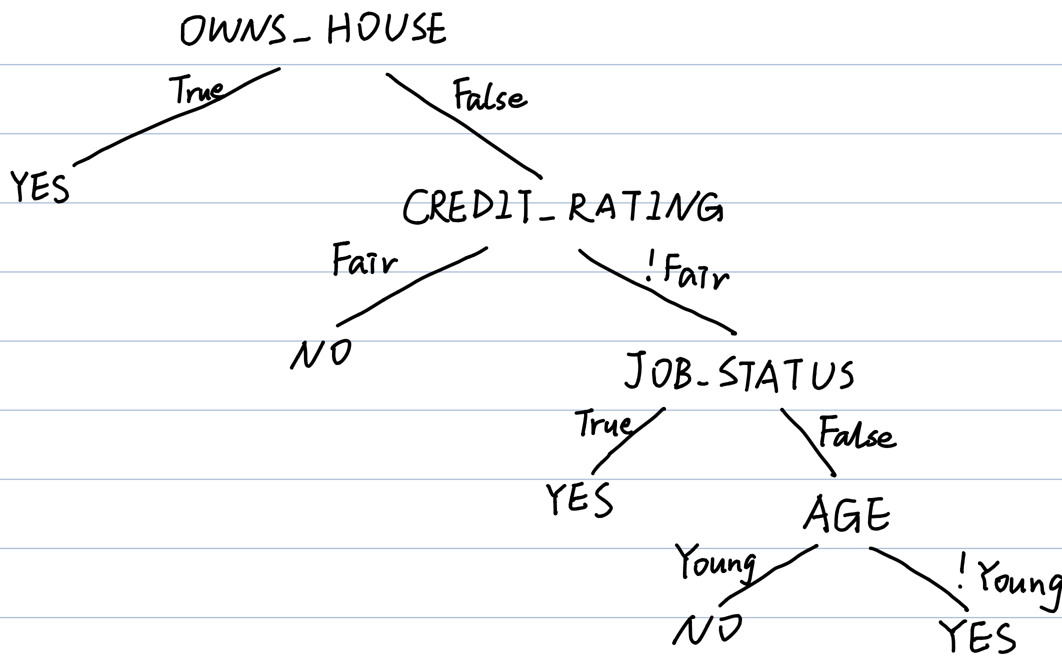
$$Gini(D, C) = \frac{6}{15}(1-1-0) + \frac{9}{10}\left(1-\left(\frac{3}{9}\right)^2-\left(\frac{6}{9}\right)^2\right) = 0.267$$

$$Gini(D, E=1) = \frac{5}{15}\left(1-\left(\frac{1}{5}\right)^2-\left(\frac{4}{5}\right)^2\right) + \frac{10}{15}\left(1-\left(\frac{8}{10}\right)^2-\left(\frac{2}{10}\right)^2\right)$$
$$= 0.32$$

$$Gini(D, E=2) = \frac{6}{15}\left(1-\left(\frac{4}{6}\right)^2-\left(\frac{2}{6}\right)^2\right) + \frac{9}{15}\left(1-\left(\frac{5}{9}\right)^2-\left(\frac{4}{9}\right)^2\right)$$
$$= 0.474$$

$$Gini(D, E=3) = \frac{4}{15}(1-1-0) + \frac{11}{15}\left(1-\left(\frac{5}{11}\right)^2-\left(\frac{6}{11}\right)^2\right)$$
$$= 0.364$$

$\therefore E = 1$




6. 有如下表所示的数据集

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_n$ | 0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.45 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 0.95 | 1.0 |
| $y_n$ | 4 | 2.4 | 1.5 | 1.0 | 1.2 | 1.5 | 1.8 | 2.6 | 3.0 | 4.0 | 4.5 | 5.0 | 6.0 |

用误差平方准则下的 CART 回归树算法生成一棵回归树，并计算当 $x = 0.76$ 时，回归函数的输出（取门限为 $\varepsilon_0 = 0.1$）。

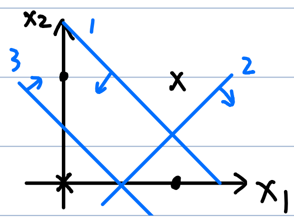
编写代码计算 CART 回归树，得到的判决门限为：

pivot = [0, 0.1, 0.5, 0.7, 0.8, 0.95]

$x = 0.76$ 时，$0.7 < x \leq 0.8$

$\therefore \hat{y} = g(0.8) = 4$

1. 对异或样本集 $\boldsymbol{D} = \left\{\left((0,0)^{\mathrm{T}},-1\right),\left((0,1)^{\mathrm{T}},1\right),\left((1,0)^{\mathrm{T}},1\right),\left((1,1)^{\mathrm{T}},-1\right)\right\}$，使用 AdaBoost 算法和类似例 8.2.1 的弱分类器~~（一个弱分类器只选择一个水平线或垂直线作为分界线）~~，给出一种集成学习器，可正确分类异或样本。




$D_1(n) = \frac{1}{4}$, $n = 1, 2, 3, 4$

选择弱分类器 $h_1$：$x_1 + x_2 = 1.5$，下方为正例

$h_1(x_n) = \{1, 1, 1, -1\}$

$\varepsilon_1 = P_{n\sim D_1(n)}(h_1(x_n) \neq y_n) = 0.25$

$\alpha_1 = \frac{1}{2}\ln\left(\frac{1-\varepsilon_1}{\varepsilon_1}\right) = 0.5493$

$$D_2(n) = \frac{D_1(n)}{Z_1}\exp(-\alpha_1 y_n h_1(x_n))$$
$$= \frac{1}{4Z_1}\{e^{\alpha_1}, e^{-\alpha_1}, e^{-\alpha_1}, e^{-\alpha_1}\}$$
$$= \{0.5, 0.167, 0.167, 0.167\}$$

其中，$Z_1 = 1.1547$

选择弱分类器 $h_2: x_1 - x_2 = 0.5$，下方为正例

$h_2(x_n) = \{-1, -1, 1, -1\}$

$\varepsilon_2 = P_{n\sim D_2(n)}(h_2(x_n) \neq y_n) = 0.167$

$\alpha_2 = \frac{1}{2}\ln\left(\frac{1-\varepsilon_2}{\varepsilon_2}\right) = 0.8047$

$$D_3(n) = \frac{D_2(n)}{Z_2}\exp(-\alpha_2 y_n h_2(x_n))$$
$$= \frac{1}{4Z_2}\{e^{-\alpha_2}, e^{\alpha_2}, e^{-\alpha_2}, e^{-\alpha_2}\}$$
$$= \{0.125, 0.625, 0.125, 0.125\}$$

其中，$Z_2 = 1.1180$

选择弱分类器 $h_3: x_1 + x_2 = 0.5$，上方为正例

$h_3(x_n) = \{-1, 1, 1, 1\}$

$\varepsilon_3 = P_{n\sim D_3(n)}(h_3(x_n) \neq y_n) = 0.125$

$\alpha_3 = \frac{1}{2}\ln\left(\frac{1-\varepsilon_3}{\varepsilon_3}\right) = 0.9730$

$H(x_n) = \mathrm{sgn}\left(\sum_{t=1}^{3}\alpha_t h_t(x)\right) = \{-1, 1, 1, -1\}$

$\therefore H(x) = \mathrm{sgn}\left(\sum_{t=1}^{3}\alpha_t h_t(x)\right)$

其中 $\alpha_1 = 0.5493$，$h_1(x) = x_1 + x_2 = 1.5$

$\alpha_2 = 0.8047$，$h_2(x) = x_1 - x_2 = 0.5$

$\alpha_3 = 0.9730$，$h_3(x) = x_1 + x_2 = 0.5$

3. 在梯度提升决策树中，若针对回归问题的一种 Huber 目标函数

$$L(y,F(\boldsymbol{x}))=\begin{cases}[y-F(\boldsymbol{x})]^2, & |y-F(\boldsymbol{x})|\le\delta \\ 2\delta|y-F(\boldsymbol{x})|-\delta^2, & 其他\end{cases}$$

这里$\delta$是给定常数，利用式（8.3.13）计算$r_{tn}=-g_{tn}$

$$r_{tn}=-g_{tn}=-\left(\frac{\partial L(y,F(x))}{\partial F(x)}\right)_{F(x_n)=F_{t-1}(x_n)}$$

$$L(y_n,F(x_n))=\begin{cases}(y_n-F(x_n))^2 \\ 2\delta\,|y_n-F(x_n)|-\delta^2\end{cases}=\begin{cases}(y_n-F(x_n))^2 & |y_n-F(x_n)|\le\delta \\ 2\delta(y_n-F(x_n)) & y_n-F(x_n)>\delta \\ -2\delta(y_n-F(x_n)) & F(x_n)-y_n>\delta\end{cases}$$

$$\frac{\partial L}{\partial F(x_n)}=\begin{cases}-2(y_n-F(x_n)) & |y_n-F(x_n)|\le\delta \\ -2\delta & y_n-F(x_n)>\delta \\ 2\delta & F(x_n)-y_n>\delta\end{cases}$$

$$r_{tn}=-g_{tn}=-\frac{\partial L}{\partial F(x_n)}=\begin{cases}2(y_n-F(x_n)) & |y_n-F(x_n)|\le\delta \\ 2\delta & y_n-F(x_n)>\delta \\ -2\delta & F(x_n)-y>\delta\end{cases}$$

4. 讨论多分类情况的梯度。设有 $K$ 种类型，任意样本 $n$ 的类型标注是 $K$-to-1 编码 $y_{ni}$。同时实现 $K$ 个提升回归函数 $F_i(\boldsymbol{x}), i=1,2,\cdots,K$，在第 $t$ 轮各类型的输出概率为 softmax 函数，

$$\hat{y}_{ni}=\frac{e^{F_{t-1,i}(x_n)}}{\sum_{k=1}^{K}e^{F_{t-1,k}(x_n)}}$$

证明：对每个提升函数 $F_i(\boldsymbol{x}), i=1,2,\cdots,K$，在第 $t$ 轮的梯度残差值为

$$r_{tni}=-g_{tni}=-\left[\frac{\partial L(y_n,F_i(\boldsymbol{x}_n))}{\partial F_i(\boldsymbol{x}_n)}\right]_{F_i(\boldsymbol{x}_n)=F_{t-1,i}(\boldsymbol{x}_n)}=y_{ni}-\hat{y}_{ni}$$

$$\hat{y}_{ni}=\frac{e^{F_{t-1,i}(x_n)}}{\sum_{K=1}^{K}e^{F_{t-1,K}(x_n)}}$$

$$L(y_n, F_i(x_n)) = -\sum_{n=1}^{N}\sum_{i=1}^{k} y_{ni} \ln \hat{y}_{ni}$$

$$\frac{\partial L}{\partial F_i(x_n)} = -\frac{y_{ni}}{\hat{y}_{ni}} \frac{\partial \hat{y}_{ni}}{\partial F_i(x_n)}$$

$$= -\frac{y_{ni}}{\hat{y}_{ni}} \cdot \frac{e^{F_i(x_n)} \sum_{k=1}^{k} e^{F_k(x_n)} - e^{F_i(x_n)} e^{F_i(x_n)}}{\left(\sum_{k=1}^{k} e^{F_k(x_n)}\right)^2}$$

$$= -y_{ni} \frac{\sum_{k=1}^{k} e^{F_k(x_n)} - e^{F_i(x_n)}}{\sum_{k=1}^{k} e^{F_k(x_n)}}$$

$$= -y_{ni}(1 - \hat{y}_{ni})$$

$$= \hat{y}_{ni} - y_{ni}$$

$$r_{tni} = -\frac{\partial L}{\partial F_i(x_n)} = y_{ni} - \hat{y}_{ni}$$