1.1. №1

$$H(X) = \sum_{k=1}^{K} p_k(1-p_k), \text{ где } p_k = \frac{\sum_{x_i \in X} I\{y_i = k\}}{|X|}$$

Случайный классификатор $y(x) = j \propto$ дискретно с p_k

P(ошибки) $\overset{?}{=} H(X)$

<u>Реш-е</u>. $P(\text{ошибки}) = P(\hat{y} \neq y) = \sum_{k=1}^{n} P(y=k, \hat{y} \neq k) = \sum_{k=1}^{n} P(y=k) \cdot P(\hat{y} \neq k) = \sum_{k=1}^{n} p_k(1-p_k)$,

т.к. мы оценили $P(y=k)$ как p_k. А вообще нам не дано $P(y=k)$, так что ⌐(!!)⌐.

1.1 №2

a) $L(y,c) = (y-c)^2 \Rightarrow \frac{1}{n}\sum_{i=1}^{n} L(Y_i, c) \longrightarrow \min_c \Rightarrow$

$\Rightarrow -\frac{2}{n}\sum_{i=1}^{n}(Y_i - c) = 0$ (т.к. параболы - строго выпуклые, сумма с полож. коэф-ми мин. тоже выпукла $\Rightarrow \exists!\ \min$). Тогда $c = \overline{Y} \Rightarrow$

$$\Rightarrow H(X) = \frac{1}{|X|} \sum_{x_i \in X} (y_i - \overline{y})^2$$

б) $\frac{1}{n} \sum_{i=1}^{n} |y_i - c| \to \min_c$

$-\frac{1}{n} \sum_{i=1}^{n} \text{sign}(y_i - c) = 0$ (опять же в силу выпуклости реш-е сущ-т и ед-но)

$\Rightarrow c = med\{y_1, .., y_n\}$.

Какое оптимальное значение в листе? Т.к. значение в листе - какая-то константа, нам хочется минимизировать его функцию ~~потерь~~ риска, которая есть $H(X) \Rightarrow$ оптимальные значения в листьях: а) $\hat{y} = \frac{1}{|X_m|} \sum_{x_i \in X_m} y_i$

б) $\hat{y} = med\{y_1, .., y_m\}$.

1.1 с/з

$$H(X) = \min_{\substack{p_1, .., p_k \in [0;1] \\ \sum_{i=1}^{k} p_i = 1}} \frac{1}{n} \sum_{i=1}^{n} L(Y_i, \{p_k\})$$

Реш-е:

а) $L(y, \{p_k\}) = \sum_{k=1}^{K} (p_k - I\{y = k\})^2$

Тогда $\frac{1}{n} \sum_{i=1}^{n} L(y_i, \{p_k\}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} (p_k^2 - 2\, I\{y_i = k\} p_k$

$+ I\{y_i = k\}) = \sum_{k=1}^{K} (p_k^2 - 2 \cdot \underbrace{\hat{p}_k}_{\frac{1}{n}\sum_{i=1}^{n} I\{y_i = k\}} p_k + \hat{p}_k) =$

$$= \sum_{k=1}^{K} p_k^2 - 2\sum_{k=1}^{K} \hat{p}_k p_k + 1 \longrightarrow \min_{\substack{\{p_k\} \\ \sum p_k = 1}}$$

$$L(\vec{p}, \lambda) = \sum_{k=1}^{K} p_k^2 - 2\sum_{k=1}^{K} \hat{p}_k p_k + 1 + \lambda\left(\sum_{k=1}^{K} p_k - 1\right)$$

Необх. усл-е ККТ (усл-е огр-е – аффинная ф-я):

$$\frac{\partial L}{\partial p_k} = 2p_k - 2\cdot\hat{p}_k + \lambda = 0 \; \cancel{(p_k - 2 + \lambda = 0)} \; \forall k$$

Зам-м, что отсюда следует: $\sum_{k=1}^{K} \frac{\partial L}{\partial p_k} = 2 - 2 + k\lambda = 0$

$\Rightarrow \lambda = 0 \Rightarrow p_k = \hat{p}_k \; \forall k$, причем $p_k \in [0;1]$.

Зам-м, что $H(X)$ выпукла как сумма выпуклых с коэф-ми $>0 \Rightarrow$ необх. усл-е явл. достаточным $\Rightarrow H(X) = \frac{1}{n}\sum_{i=1}^{n} L(y_i, \{\hat{p}_k\}) =$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(\hat{p}_k - I\{y_i = k\}\right)^2 = 1 - \sum_{k=1}^{K}\hat{p}_k^2 = \sum_{k=1}^{K}\hat{p}_k(1-\hat{p}_k)$$

Тогда и в листьях будут такие оптимальные вер-ти: $p_k = \hat{p}_k$.

б) $L(y, \{p_k\}) = -\sum_{k=1}^{K} I\{y = k\}\log p_k \rightarrow$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} L(y_i, \{p_k\}) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K} I\{y_i = k\}\log p_k =$$

$$= -\sum_{k=1}^{K}\hat{p}_k \log p_k \longrightarrow \min_{\sum_{k=1}^{K} p_k = 1}$$

$$L(\vec{p}, \lambda) = -\sum_{k=1}^{K}\hat{p}_k\log p_k + \lambda\left(\sum_{k=1}^{K} p_k - 1\right)$$

Т.к. $\hat{p}_k \geq 0$, то $-\sum_{k=1}^{K}\hat{p}_k\log p_k$ выпукла, ограничения линейны $\Rightarrow$ необх. усл-е (ККТ) явл. достаточными:

$$\frac{\partial L}{\partial p_k} = -\frac{\hat{p}_k}{p_k} + \lambda = 0 \Rightarrow \cancel{\sum_{k=1}^{K} \frac{\partial L}{\partial p_k}}$$

$$\Rightarrow \lambda p_k = \hat{p}_k \Rightarrow \lambda \sum_{k=1}^{K} p_k = \sum_{k=1}^{K} \hat{p}_k \Rightarrow \lambda = 1 \Rightarrow$$

$$\Rightarrow p_k = \hat{p}_k \in [0;1] \Rightarrow H(X) = -\sum_{k=1}^{K} \hat{p}_k \log \hat{p}_k$$

Причём оптимальные значения вер-тей классов в листьях: $p_k = \hat{p}_k$.

1.1 а) 4

\# Выборка: n, признаки: d, глубина: D. $T(H(x)) = O(1)$

<u>Реш-е.</u> На каждом шаге мы делаем след. действия: для каждого из d признаков ~~мы~~ перебираем все возможны $|X_m|$ значений и считаем $H(X)$, (число объектов в вершине) потом $Q(X, j, t)$. Т.к. на каждом из уровней всегда n объектов, то суммарно переберётся $n \cdot d \cdot D$

Тогда в итоге: $O(nd \cdot D)$