Домашнее задание №5.

№1. $X \geq 0$ — СВ, $\forall \varepsilon > 0: P(X \geq \varepsilon) \leq e^{-2m\varepsilon^2}$

Д-ть, что $E\{e^{2(m-1)X^2}\} \leq m$

◇:

$$P(X \geq \varepsilon) = 1 - P(X < \varepsilon) \leq e^{-2m\varepsilon^2}$$

$$F_X(\varepsilon) \geq 1 - e^{-2m\varepsilon^2}$$

$$E\{e^{2(m-1)X^2}\} = \int_0^{+\infty} e^{2(m-1)t^2} p_X(t)dt = \int_0^{+\infty} e^{2(m-1)t^2} dF_X(t)$$

$$\int_0^a e^{2(m-1)t^2} dF_X(t) = \left[\begin{matrix} u = e^{2(m-1)t^2} & du = 4(m-1)te^{2(m-1)t^2} \\ dv = dF_X(t) & v = F_X(t) \end{matrix}\right] =$$

$$= e^{2(m-1)a^2} F_X(a) - \int_0^a 4(m-1)te^{2(m-1)t^2} \cdot F_X(t)dt \leq$$

$$\leq e^{2(m-1)a^2} \cdot F_X(a) - \int_0^a 4(m-1)te^{2(m-1)t^2}(1 - e^{-2mt^2})dt =$$

$$= e^{2(m-1)a^2} \cdot F_X(a) - \int_0^a 4(m-1)te^{2(m-1)t^2}dt + \int_0^a 4(m-1)te^{2(m-1)t^2}e^{-2mt^2}dt =$$

$$= e^{2(m-1)a^2} \cdot F_X(a) - \int_0^a e^{2(m-1)t^2} d(2(m-1)t^2) + \int_0^a 4(m-1)te^{-2t^2}dt =$$

$$= e^{2(m-1)a^2} F_X(a) - (e^{2(m-1)a^2} - 1) + (m-1)\int_0^a e^{-2t^2} d(2t^2) =$$

$$= 1 - e^{2(m-1)a^2} \cdot (1 - F_X(a)) + (m-1) \cdot (1 - e^{-2t^2})$$

При $a \to +\infty$

$$E\{e^{2(m-1)X^2}\} \leq 1 + (1 - e^{-2t^2})(m-1) \leq 1 + m - 1 = m.$$

⊠

N2. H - конечный класс гипотез P - равномерное распр. над H

Q: $Q(h_S) = 1$, h_S
$Q(h) = 0$, $h \in H$, $h \neq h_S$.

①. Д: $L_D(h_S) \leq L_S(h_S) + \sqrt{\dfrac{\ln(|H|) + \ln(\frac{m}{\delta})}{2(m-1)}}$

Воспользуемся теоремой для PAC-Байес из лекции.

$$L_D(Q) \leq L_S(Q) + \sqrt{\frac{D(Q\|P) + \ln(\frac{m}{\delta})}{2(m-1)}}$$

Т.к. P - равномерно распределено над классом H (конечным), то $P(h) = \dfrac{1}{|H|}$.

$$D(Q\|P) = E_{h \sim Q}\left\{\ln\left(\frac{Q(h)}{P(h)}\right)\right\} = \sum_{h \in H} \ln\frac{Q(h)}{P(h)} Q(h) = \left[\begin{array}{l} Q(h_S)=1 \\ Q(h)=0 \\ h \neq h_S \end{array}\right] =$$

$$= \ln\frac{Q(h_S)}{P(h_S)} \cdot Q(h_S) = \ln\frac{1}{P(h_S)} = \ln(|H|).$$

Т.к. $Q(h_S) = 1$, то $L_D(h_S) = L_D(Q)$ и $L_S(h_S) = L_S(Q)$.

$\Rightarrow$ Получим, что
$$L_D(h_S) \leq L_S(h_S) + \sqrt{\frac{\ln(|H|) + \ln(\frac{m}{\delta})}{2(m-1)}}$$

$\otimes$

②. Бритва Оккама:
Краткое объяснение (короткая гипотеза) скорее всего лучше, чем длинное.
Сравним нер-ва из теорем и для бритвы Оккама:

$$L_D(h) \leq L_S(h) + \sqrt{\frac{|h| + \ln\frac{2}{\delta}}{2m}} \text{ и } L_D(Q) \leq L_S(Q) + \sqrt{\frac{\ln(|H|) + \ln(\frac{m}{\delta})}{2(m-1)}}.$$

Аналогично можно сформулировать следующее правило:

Меньший конечный класс гипотез скорее всего лучше, чем больший.

Как видно из неравенства, при одинаковом эмпирическом риске оценка на true risk будет меньше у гипотезы из класса с меньшей мощностью.