Статистическое обучение
Задание 1

Сергей Володин, 374 гр.

задано 2017.02.19

Meta

Делал один. Список ссылок:

1. http://www.stat.cmu.edu/~arinaldo/36788/subgaussians.pdf
2. https://en.wikipedia.org/wiki/Holder's_inequality

Упражнение 1

1. Неравенство Маркова: Если $X \geqslant 0$, то $P(X \geqslant \varepsilon) \leqslant \frac{\mathbb{E}X}{\varepsilon}$. Нужно: $P(X \geqslant \varepsilon) = \frac{\mathbb{E}X}{\varepsilon}$. Найдем $P(X < \varepsilon) = 1 - \frac{\mathbb{E}X}{\varepsilon}$, $f_X(x) = \frac{\mathbb{E}X}{x^2}$. Тогда $\mathbb{E}X = \int\limits_0^\infty x f_X(x)dx = \int\limits_0^\infty \mathbb{E}X \frac{dx}{x}$. Поскольку интеграл $\int\limits_0^\infty \frac{dx}{x}$ расходится, то $\mathbb{E}X = 0$. Значит, $\boxed{X = 0}$. Проверим: $0 = P(0 \geqslant \varepsilon) = \frac{0}{\varepsilon}$ ■
2. Неравенство Чебышева: $P(|X - \mathbb{E}X| \geqslant a) \leqslant \frac{\sigma^2}{a^2}$. Если обозначить $\eta = |X - \mathbb{E}X|^2$, то получим неравенство Маркова. Возьмем предыдущий пример $\Rightarrow \eta = 0 \Rightarrow X = c$ (константа). Проверим: $0 = P(0 \geqslant a) = \frac{0}{a^2}$ (для константы $\sigma = 0$) ■

Упражнение 2.1

Имеем: $Y \geqslant 0$ — случайная величина, числа $A \geqslant 2$, $B > 0$. $\forall \varepsilon \geqslant 0 \hookrightarrow P(Y \geqslant \varepsilon) \leqslant A\exp(-\frac{\varepsilon^2}{B^2})$.

1. Оценим $\mathbb{E}e^{\lambda Y^2} = 1 + \int\limits_1^\infty P(e^{\lambda Y^2} > x)dx$. Перепишем $e^{\lambda Y^2} > x \Leftrightarrow \lambda Y^2 > \ln x \Leftrightarrow Y > \sqrt{\frac{\ln x}{\lambda}}$. Значит, $\mathbb{E}e^{\lambda Y^2} \leqslant 1 + A\int\limits_1^\infty x^{-1/\lambda B^2}dx = 1 + A\frac{1}{1/\lambda B^2 - 1}$ при условии $\lambda \in (0, 1/B^2)$. Берём $\lambda = 1/2B^2$. Тогда $\mathbb{E}e^{\lambda Y^2} \leqslant 1 + A \leqslant 2A$ при $A \geqslant 2$
2. $\mathbb{E}Y = \sqrt{\frac{1}{\lambda}\ln e^{\lambda(\mathbb{E}Y)^2}} \underbrace{\leqslant}_{\text{Йенс. } e^{\lambda x^2}} \sqrt{\frac{1}{\lambda}\ln \mathbb{E}e^{\lambda Y^2}} \underbrace{\leqslant}_{(1)} \sqrt{2B^2\ln 2A} = \sqrt{2}B\sqrt{\ln 2A}$. Заметим, что при $A \geqslant 2$, $\sqrt{\ln 2A} \leqslant \sqrt{2\ln A}$. Тогда $\mathbb{E}Y \leqslant \boxed{2B\sqrt{\ln A}}$. То есть, проведено доказательство для $C = 2$.

Упражнение 2.2

Имеем: $Y \geqslant 0$ — случайная величина, числа $A \geqslant 2$, $B > 0$. $\forall \varepsilon \geqslant 0 \hookrightarrow P(Y \geqslant \varepsilon) \leqslant A\exp(-\frac{\varepsilon}{B})$.

1. Оценим $\mathbb{E}e^{\lambda Y} = 1 + \int\limits_1^\infty P(e^{\lambda Y} > x)dx$. Рассмотрим $e^{\lambda Y} > x \Leftrightarrow Y > \frac{\ln x}{\lambda}$. $P(Y > \frac{\ln x}{\lambda}) \leqslant Ae^{-\frac{\ln x}{\lambda B}} = Ax^{-1/\lambda B}$. Тогда $\mathbb{E}e^{\lambda Y} \leqslant 1 + A\int\limits_1^\infty x^{-1/\lambda B}dx$ при $\lambda B < 1$. Берем $\lambda = 1/2B$. Тогда $\mathbb{E}e^{\lambda Y} \leqslant 1 + A \leqslant 2A$
2. $\mathbb{E}Y = \frac{1}{\lambda}\ln e^{\lambda \mathbb{E}Y} \leqslant \frac{1}{\lambda}\ln \mathbb{E}e^{\lambda Y} \leqslant \frac{1}{\lambda}2A = 2B\ln 2A \leqslant \boxed{4B\ln A}$

Упражнение 3.1

http://www.stat.cmu.edu/~arinaldo/36788/subgaussians.pdf
https://en.wikipedia.org/wiki/Holder's_inequality
Случайная величина X — субгауссовская с параметром $\sigma \Leftrightarrow \mathbb{E}e^{\lambda X} \leqslant e^{\frac{\lambda^2\sigma^2}{2}}$.
Пусть X_1, X_2 — субгауссовские с параметрами σ_1 и σ_2. $Y = X_1 + X_2$. Доказать: Y — субгауссовская для некоторого σ.
$\mathbb{E}e^{\lambda Y} = \mathbb{E}e^{\lambda X_1}e^{\lambda X_2}$.
Неравенство Гёльдера для мат.ожиданий $\xi, \eta \geqslant 0$, $1/p + 1/q = 1$:
$\mathbb{E}\xi\eta \leqslant (\mathbb{E}\xi^p)^{1/p}(\mathbb{E}\eta^q)^{1/q}$
Тогда $\mathbb{E}e^{\lambda Y} \leqslant (\mathbb{E}e^{p\lambda X_1})^{1/p}(\mathbb{E}e^{q\lambda X_2})^{1/q} \leqslant (e^{(p\lambda)^2\sigma_1^2/2})^{1/p}(e^{(q\lambda)^2\sigma_2^2/2})^{1/q} = e^{\frac{\lambda^2}{2}(p\sigma_1^2 + q\sigma_2^2)} \to \min\limits_{1/p+1/q=1}$

Поскольку $1/p+1/q=1$, $q=\frac{p}{p-1}$. Тогда $p\sigma_1^2+q\sigma_2^2=p\sigma_1^2+\frac{p}{p-1}\sigma_2^2\to\min\limits_p \Rightarrow p=\frac{\sigma_1+\sigma_2}{\sigma_1}$, $q=\frac{\sigma_1+\sigma_2}{\sigma_2}$.
Тогда $\mathbb{E}e^{\lambda Y}\leqslant e^{\frac{\lambda^2(\sigma_1+\sigma_2)^2}{2}}$
Тогда $\boxed{\sigma=\sigma_1+\sigma_2}$

Упражнение 3.1.1

Пусть $X_1=X_2\sim N(0,\sigma_0^2)$. Тогда $\mathbb{E}e^{\lambda X_i}=e^{\lambda^2\sigma_0^2/2}$. А $\mathbb{E}e^{\lambda(X_1+X_2)}=\mathbb{E}e^{2\lambda X_i}=e^{\lambda^2(2\sigma_0)^2/2}$. В этом примере $\sigma=2\sigma_0$ ■

Упражнение 3.2

Пусть $X\sim N(0,1)$. Тогда X — субгауссовская с параметром 1. Определим $X_1=X$, $X_2=X$ — две субгауссовские величины. Определим $Y=X_1X_2$. Но $\mathbb{E}Y=\mathbb{E}X^2=1\neq 0$, значит, Y не может быть субгауссовской

Упражнение 3.3

Обозначим $f(\lambda)=\mathbb{E}e^{\lambda X}$, $g(\lambda)=e^{\sigma^2\lambda^2/2}$. $f(\lambda)=\sum\limits_{k=0}^{\infty}\frac{\lambda^k}{k!}\mathbb{E}X^k=1+\lambda\mathbb{E}X+\frac{\lambda^2}{2}\mathbb{E}X^2+\dots$. Значит, $f'(0)=\mathbb{E}X$. Найдём $g'(\lambda)=\lambda g(\lambda)$. Найдём $f(0)=g(0)=1$. Значит, $g'(0)=0$. Обозначим $h(\lambda)=f(\lambda)-g(\lambda)$. По условию, $h(\lambda)\leqslant 0$. Поскольку $h(0)=0$, то $h'(0)=0$. Но $h'(0)=f'(0)-g'(0)=\mathbb{E}X$. Значит, $\boxed{\mathbb{E}X=0}$

Упражнение 3.4

Source: *http://www.stat.cmu.edu/~arinaldo/36788/subgaussians.pdf*
Пусть $\xi\geqslant 0$ — случайная величина, $f(\xi)$ — функция: $f(0)=0$. Докажем, что $\mathbb{E}f(\xi)=\int\limits_0^\infty f'(t)P(\xi>t)dt$

$$\int\limits_0^\infty f'(t)P(\xi>t)dt=\int\limits_0^\infty dt f'(t)\int\limits_t^\infty dq f_\xi(q)=\int\limits_0^\infty dq f_\xi(q)\underbrace{\int\limits_0^q f'(t)dt}_{f(q)-\cancel{f(0)}}=\int\limits_0^\infty f_\xi(q)f(q)dq=\mathbb{E}f(\xi)\ ■$$

Тогда $\mathbb{E}|X|^p=\big|f(y)=y^p,\ \xi=|X|\big|=\int\limits_0^\infty py^{p-1}P(|X|>y)dy$
Поскольку X — субгауссовская с параметром σ, то по неравенству Чернова $P(|X|>y)\leqslant 2e^{-y^2/2\sigma^2}$. Подставим в интеграл:
$\mathbb{E}|X|^p\leqslant\int\limits_0^\infty py^{p-1}2e^{-y^2/2\sigma^2}dy=\left|t=\frac{y^2}{2\sigma^2},\ dy=\sqrt{\frac{\sigma^2}{2t}}dt\right|=\int\limits_0^\infty p(2t\sigma^2)^{(p-1)/2}2e^{-t}\sqrt{\sigma^2/2t}dt=p(2\sigma^2)^{p/2}\int\limits_0^\infty t^{p/2-1}e^{-t}dt=$
$=p(2\sigma^2)^{p/2}\Gamma(p/2)$.
Поскольку $\Gamma(n+1)=n!\sim\sqrt{2\pi n}(\frac{n}{e})^n$ (Формула Стирлинга).
Тогда $\mathbb{E}|X|^p\leqslant C(\sigma)p(2\sigma^2)^{p/2}\sqrt{\pi p}(\frac{p}{2e})^{p/2}$ и $C(\sigma)\geqslant 1$
Значит, $(\mathbb{E}|X|^p)^{p/2}\leqslant C^{1/p}(\sigma)p^{3/2p}(2\sigma^2)^{1/2}\sqrt{\pi}(2e)^{-1/2}p^{1/2}$
Поскольку $C\geqslant 1$ и $p\geqslant 1$, $C^{1/p}\leqslant p$
$p^{3/2p}\leqslant D$ — ограниченная функция
Получаем $(\mathbb{E}|X|^p)^{p/2}\leqslant\underbrace{C(\sigma)D(2\sigma^2)^{1/2}\sqrt{pi/2e}}_{K(\sigma)}p^{1/2}=K(\sigma)\sqrt{p}$ ■

Упражнение 4.1

Плотность нормального распределения: $\psi(x)=f_{N(0,1)}(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. Тогда $\frac{d}{dx}\psi(x)=\frac{1}{\sqrt{2\pi}}(-2x/2)e^{-\frac{x^2}{2}}=-x\psi(x)$.
Значит, $\boxed{x\psi(x)+\psi'(x)=0}$

Упражнение 4.2

Обозначим $f_1(x)=\psi(x)(1/x-1/x^3)$, $f_2(x)=P(X\geqslant x)=\int\limits_x^\infty\psi(t)dt$, $f_3(x)=\psi(x)(1/x-1/x^3+3/x^5)$. Доказать: при $x>0$ $f_1\leqslant f_2\leqslant f_3$. Обозначим $g(x)=f_2(x)-f_1(x)$, $h(x)=f_3(x)-f_2(x)$ Нужно доказать, что $g,h\geqslant 0$.
Тогда $f_1'(x)=-x\psi(x)(1/x-1/x^3)+\psi(x)(-1/x^2+3/x^4)=\psi(x)(3/x^4-1)$, $f_2'(x)=-\psi(x)$, $f_3'(x)=\psi(x)(-15/x^6-1)$.
Тогда $g'(x)=-\frac{3\psi(x)}{x^4}<0$, $h'(x)=-\frac{15\psi(x)}{x^6}<0$.
$g(+0)=+\infty$, $h(+0)=+\infty$.

$$g(+\infty)=\lim_{x\to\infty}\cancel{f_2(x)}^{0}-f_1(x)=-\lim_{x\to\infty}\underbrace{\psi(x)}_{\to 0}\underbrace{(1/x-1/x^3)}_{\to 0}=0$$

$$h(+\infty)=\lim_{x\to\infty}f_3(x)-\underbrace{f_2(x)}_{\to 0}=\lim_{x\to\infty}\underbrace{\psi(x)}_{\to 0}\underbrace{(1/x-1/x^3+3/x^5)}_{\to 0}=0$$

Получаем две строго монотонно убывающие непрерывные функции g, h на $(0, +\infty)$, причем обе стремятся к 0. Значит, $\forall x > 0 \hookrightarrow g, h > 0$ ■

Упражнение 4.3

Рассмотрим $P(X \geqslant x) \leqslant \inf\limits_{\lambda>0} \frac{\mathbb{E}e^{\lambda X}}{e^{\lambda x}}$. Для нормальной случайной величины $\mathbb{E}e^{\lambda X} = e^{\frac{\lambda^2\sigma^2}{2}+\mu x} = e^{\frac{\lambda^2}{2}}$. Значит, $\inf\limits_{\lambda>0} \frac{\mathbb{E}e^{\lambda X}}{e^{\lambda x}} = \inf\limits_{\lambda>0} \exp(\frac{\lambda^2}{2} - \lambda x)$. Находим $\lambda^* = x > 0$, получаем $P(X \geqslant x) \leqslant \exp(-\frac{x^2}{2}) = \sqrt{2\pi}\psi(x)$

Имеем две оценки: $\begin{cases} P(X \geqslant x) \leqslant \psi(x)(1/x - 1/x^3 + 3/x^5) \\ P(X \geqslant x) \leqslant \sqrt{2\pi}\psi(x) \end{cases}$

Поделим $\frac{\psi(x)(1/x-1/x^3+3/x^5)}{\sqrt{2\pi}\psi(x)} = \frac{1}{\sqrt{2\pi}}(1/x - 1/x^3 + 3/x^5) \sim \frac{1}{\sqrt{2\pi}x}$. Значит, оценка в (4.2) лучше, чем оценка в (4.3).

Упражнение 4.4

Рассмотрим $\xi = \sum\limits_{i=1}^{n} \xi_i$, $\{\xi_i\}$ — i.i.d., $\xi_i \sim Be(p)$:

$$\xi_i = \begin{cases} 1, & p \\ 0, & 1-p \end{cases}$$

$q = 1-p$. $\zeta_i = \frac{\xi_i - p}{\sqrt{npq}}$. Тогда $\mathbb{E}\zeta_i = 0$. Обозначим $S = \sum\limits_{i=1}^{n} \zeta_i$. Тогда $\mathbb{E}S = 0$ $\mathbb{D}S = 1$.

$\zeta_i \in [-\frac{p}{\sqrt{npq}}, \frac{1-p}{\sqrt{npq}}]$ — субгауссовская с $\sigma_i^2 = \frac{1}{4npq}$ по Лемме Хёффдинга.

1. Неравенство Хёффдинга.
$$P(S \geqslant \varepsilon) \leqslant e^{-\varepsilon^2/2n\sigma_i^2} = e^{-2\varepsilon^2 pq}$$

2. Теорема Муавра-Лапласа:
$$P(\sum \xi_i \geqslant \varepsilon) \approx \frac{1}{\sqrt{2\pi np(1-p)}} \int\limits_{\varepsilon}^{\infty} e^{-\frac{1}{2}\left(\frac{x-np}{\sqrt{npq}}\right)^2} dx$$

Найдем $P(S \geqslant \varepsilon) = P(\frac{\sum \xi_i - np}{\sqrt{npq}} \geqslant \varepsilon) = P(\sum \xi_i \geqslant \varepsilon\sqrt{npq} + np) = \frac{1}{\sqrt{2\pi np(1-p)}} \int\limits_{\varepsilon\sqrt{npq}}^{\infty} e^{-\frac{1}{2}\left(\frac{x}{\sqrt{npq}}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int\limits_{\varepsilon}^{\infty} e^{-x^2/2} dx$

Первая оценка $F_1(\varepsilon) = e^{-2\varepsilon^2 p(1-p)}$ зависит от p. Вторая оценка $F_2(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{\varepsilon}^{\infty} e^{-x^2/2} dx$ не зависит от p. Пусть $n \to \infty$.

1. $p \to 0$ или $q \to 0$. Тогда Неравенство Хёффдинга даст $F_1(\varepsilon) = 1$, то есть, оценка бесполезна. Теорема Муавра-Лапласа даст число $F_2(\varepsilon) < 1$. В этом случае теорема Муавра-Лапласа лучше.

2. $p = \frac{1}{2}$. Тогда $1/4 = pq \to \max$. Значения $F_1(\varepsilon) = e^{-\varepsilon^2/2}$, $F_2(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{\varepsilon}^{\infty} e^{-x^2/2} dx$. Обозначим $h(x) = F_1(x) - F_2(x)$. Тогда $h(0) = 1/2$, $h'(x) = e^{-x^2/2}(1/\sqrt{2\pi} - x)$. h возрастает, а затем убывает. $h(+\infty) = 0$. То есть, всегда $F_1 \geqslant F_2$. Это значит, что даже при $p = 1/2$ теорема Муавра-Лапласа дает лучшую оценку

Упражнение 5.1

$f^* = \arg\min\limits_{f \in Y^X} L(f)$. $L(f) = \mathbb{E}_{X\times Y}[f(X) \neq Y] = \mathbb{E}_X \mathbb{E}_{Y|X}[f(X) \neq Y] = \mathbb{E}_X P(f(X) \neq Y \big| X)$. Фиксируем X, т.е. рассмотрим одно слагаемое (или подынтегральный член):

$P(f(X) \neq Y \big| X) = P(f = 1 \big| X, Y = -1)P(Y = -1 \big| X) + P(f = -1 \big| X, Y = 1)P(Y = 1 \big| X) \boxed{=}$. Поскольку f зависит только от X, $\boxed{=} [f(X) = 1]P(Y = -1 \big| X) + [f(X) = -1]P(Y = 1 \big| X)$. В этой сумме одна из скобок $[f(X) = \cdot]$ равна 1, а другая 0, в зависимости от значения f на X. Значит, для минимизации $\mathbb{E}_{Y|X}[f(X) \neq Y]$ нужно взять $f(X) = \arg\min\limits_{j} P(Y = j \big| X)$

Рассмотрим $\eta(x) = \mathbb{E}[Y|X = x] = P(Y = 1 \big| X = x) - P(Y = -1 \big| X = x)$. Значит, $\operatorname{sign}\eta(x) = \arg\min\limits_{j} P(Y = j \big| X = x)$, то есть, $f^*(x) = \operatorname{sign}\eta(x)$ ■

Упражнение 5.2

Фиксируем x. Обозначим $p = P(Y = +1|X = x)$. Тогда $\eta(x) = P(Y = +1|X = x) - P(Y = -1|X = x) = p - (1-p) = 2p - 1$. Знаем, что $|2p - 1| \geqslant h$. Значит, либо $p \geqslant \frac{h+1}{2}$, либо $p \leqslant \frac{1-h}{2}$

Поскольку $f^* = \operatorname{sign}(2p - 1)$, то $[f^* = +1] = [p > 0.5]$, а $[f^* = -1] = [p < 0.5]$

Рассмотрим $L(f^*) = \mathbb{E}_X \left(\underbrace{[p > 0.5](1-p) + [p < 0.5]p}_{l(x)} \right)$.

1. Пусть $p > 0.5$. Но тогда $p \geqslant \frac{1+h}{2}$. Значит, $l(x) \leqslant 1 - p = \frac{1-h}{2}$
2. Пусть $p < 0.5$. Тогда $p \leqslant \frac{1-h}{2}$. Значит, $l(x) \leqslant p = \frac{1-h}{2}$

Получаем, что $l(x) \leqslant \frac{1-h}{2}$. Тогда $L(f^*) = \mathbb{E}_X l(x) \leqslant \frac{1-h}{2}$ ■

Задача 1

Имеем $Y = \{-1, 1\}$ — метки классов, K — класс функций, $f^* = \underset{f \in Y^X}{\arg\min}\, L(f)$.

Рассмотрим различные элементы $x_1, x_2, x_3 \in X$.

Определим $f_i(x) = \begin{cases} \overline{f^*(x)}, & x \neq x_i \\ f^*(x), & x = x_i \end{cases}$, где $\overline{1} = -1$, $\overline{-1} = 1$.

Определим $F = \{f^*, \overline{f^*}, f_1, f_2, f_3\}$.

1. Halving (большинство). Рассмотрим произвольный $x \in X$ (первый шаг алгоритма). Если $x = x_i$, то получим значения функций $(f^*, \overline{f^*}, \overline{f^*}, f^*, \overline{f^*})$. Halving выдаст неверный ответ (3>2), то есть, $\overline{f^*}$. Если $x \neq x_i$, то получим значения функций $(f^*, \overline{f^*}, \overline{f^*}, \overline{f^*}, \overline{f^*})$. Halving снова выдаст неверный ответ. То есть, количество ошибок Halving на произвольной выборке как минимум 1
2. Меньшинство без удалений. Алгоритм: голосуем меньшинством функций из F, не удаляем функции при неверном ответе. Пусть $x = x_i$. Получим значения $(f^*, \overline{f^*}, \overline{f^*}, f^*, \overline{f^*})$. Меньшинство: f^* (2 против 3). Получим правильный ответ. Пусть $x \neq x_i$. Получим значения $(f^*, \overline{f^*}, \overline{f^*}, \overline{f^*}, \overline{f^*})$. Меньшинство: f^* (1 против 4). Снова правильный ответ. Поскольку удалений нет, на последующих объектах также не будет ошибок.

Построен алгоритм, который для данного F делает 0 ошибок, когда Halving делает как минимум 1 ■

Задача 2

Пусть X — объекты, $Y = \{-1, 1\}$ — метки классов. Пусть $F \subseteq Y^X$ — класс функций. $\exists f^* \in F \colon Y = f^*(X)$.
Определения обучаемости F:

$1-\delta$: $\exists A_n \colon (X \times Y)^n \to Y^X\, \exists n(\varepsilon, \delta)\, \forall \varepsilon > 0\, \forall \delta \in (0,1)\, \forall N \geqslant n(\varepsilon, \delta) \hookrightarrow P(L(A(X_N)) - \inf_{f \in F} L(f) \leqslant \varepsilon) \geqslant 1 - \delta$

$\frac{1}{2}$: $\exists A_n \colon (X \times Y)^n \to Y^X\, \exists n(\varepsilon)\, \forall \varepsilon > 0\, \forall N \geqslant n(\varepsilon) \hookrightarrow P(L(A(X_N)) - \inf_{f \in F} L(f) \leqslant \varepsilon) \geqslant \frac{1}{2}$

Докажем, что $(1-\delta) \Leftrightarrow (\frac{1}{2})$

1. $(1-\delta) \Rightarrow \frac{1}{2}$: определим A_n во втором определении как A_n из первого, определим $n(\varepsilon) = n(\varepsilon, \frac{1}{2})$. Тогда

$$\forall \varepsilon > 0\, \forall N \geqslant n(\varepsilon, \frac{1}{2}) \hookrightarrow P(L(A(X_N)) - \inf_{f \in F} L(f) \leqslant \varepsilon) \geqslant \frac{1}{2}\ ■$$

2. $\frac{1}{2} \Rightarrow (1-\delta)$: Пусть $\overline{1-\delta}$ и $\frac{1}{2}$:

$$\exists A_n \colon (X \times Y)^n \to Y^X\, \exists n(\varepsilon)\, \forall \varepsilon > 0\, \forall N \geqslant n(\varepsilon) \hookrightarrow P(L(A(X_N)) - \inf_{f \in F} L(f) \leqslant \varepsilon) \geqslant \frac{1}{2}$$

$\forall A_n \colon X^n \to Y^X\, \forall n\, \exists \varepsilon > 0\, \exists \delta \in (0,1)\, \exists N \geqslant n : P(L(A(X_N)) - \inf_{f \in F} L(f) \leqslant \varepsilon) < 1 - \delta$

Поскольку $\exists f^* \in F \colon f^*(X) = Y$, то $\inf_{f \in F} L(f) = 0$

Обозначим $Q(X_N) = L(A(X_N)) = \mathbb{E}_{X \times Y}[A(X_N)(x) \neq y] = P_{X \times Y}(A(X_N)(x) \neq y) = P_X(A(X_N)(x) \neq f^*(x))$

Обозначим $\alpha_{AFL}(\varepsilon, N) = P\left(L(A(X_N)) \leqslant \varepsilon\right) = P_{X_N}(\underbrace{P_X(A(X_N)(x) \neq f^*(x))}_{L(A(X_N))} \leqslant \varepsilon) = P_{X_N}(Q(X_N) \leqslant \varepsilon)$

Тогда имеющиеся условия можно переписать:

$$\begin{cases} \exists(A, n(\varepsilon)) & \forall(\varepsilon, N \geqslant n(\varepsilon)) & \hookrightarrow & \alpha(\varepsilon, N) & \geqslant & \frac{1}{2} \\ \forall(A, n(\varepsilon, \delta)) & \exists(\varepsilon, \delta, N \geqslant n(\varepsilon, \delta)) & : & \alpha(\varepsilon, N) & < & 1 - \delta \end{cases}$$

Возьмем $(A, n(\varepsilon))$ для (2) из (1). Тогда $\exists(\varepsilon, N \geqslant n(\varepsilon), \delta) \colon \frac{1}{2} \leqslant \alpha(\varepsilon, N) < 1 - \delta$. Значит, $\delta < \frac{1}{2}$

В качестве алгоритма A выберем алгоритм из (1). Фиксируем $\varepsilon > 0$, $\delta > 0$.

Найдем $n(\varepsilon, \delta)$, такое что $\forall N \geqslant n(\varepsilon, \delta) \hookrightarrow \alpha(\varepsilon, N) \geqslant 1 - \delta$.

Выберем $N \geqslant \max\{n(t), n(\varepsilon)\}$. Величину $t < \varepsilon$ выберем позже (при выборе нового t допустимые N могут только увеличиться, что только увеличит $\alpha(\varepsilon, N)$)

Тогда $\alpha(\varepsilon, N) \geqslant \frac{1}{2}$, так как $N \geqslant n(\varepsilon)$. Также $\alpha(t, N) \geqslant \frac{1}{2}$

Рассмотрим $\alpha(\varepsilon, N) = \underbrace{P_{X_N}(Q(X_N) = 0) + P_{X_N}(Q(X_N) \in (0, t])}_{\alpha(t,N) \geqslant 1/2} + P_{X_N}(Q(X_N) \in (t, \varepsilon]) = p_0 + p_t + q_t \geqslant 1/2$.

Пусть $t \to 0$. Тогда $p_t \to 0$, $q_t \to \alpha(\varepsilon, N) - p_0$ Пусть $t \to \varepsilon$. Тогда $p_t \to \alpha(\varepsilon, N) - p_0$, $q_t \to 0$

(a) $p_0 < \delta$. Тогда $q_t \overset{t \to 0}{\longrightarrow} \alpha(\varepsilon, N) - p_0 \geqslant 1/2 - p_0 > 1/2 - \delta$. Значит, можно выбрать t, такое что $q_t \geqslant 1/2 - \delta$. После выбора t получим $N \geqslant n(t)$ и $\alpha(t, N) = p_0 + p_t \geqslant 1/2$. Значит, $\alpha(\varepsilon, N) = \underbrace{p_0 + p_t}_{\alpha(t,N) \geqslant 1/2} + \underbrace{q_t}_{\geqslant 1/2 - \delta} \geqslant 1 - \delta$

(b) $p_0 \geqslant \delta$. Тогда возможен случай $\forall N \geqslant n_0$ $P(Q(X_N) = 0) \geqslant 1/2$.