Школа анализа данных
Восстановление зависимостей
Домашнее задание №3

Кошман Дмитрий

Задача 1

Доказать:

$$P\left(\sup_x |F_n(x) - F(x)| \xrightarrow[n\to\infty]{} 0\right) = 1$$

где $F(a) = P(x < a), F_n(a) = \frac{\sum_i [x_i < a]}{L}, L-$ размер выборки

Интересующая нас сигма алгребра событий порождена множествами вида $\{x < a\}$. Найдем функцию роста относительно множества $S = \{\{x | x < a\} | a \in \mathbb{R}\}$:

$$m^S(L) = \max_{X^{(L)}} \Delta^S(X^{(L)}) = \max_{X^{(L)}} \Big(\text{мощность множества подвыборок } X^{(L)}, \text{ индуцированных } S\Big)$$

Поскольку каждая подвыборка, порожденная элементом $s \in S, s = \{x < a\}$ однозначно определяется расположением a между двумя соседними элементами выборки, то разных подвыборок не больше таких расположений, то есть $L + 1$, и это число всегда достижимо. Значит,

$$m^S(L) = L + 1$$

Теперь воспользуемся достаточным условием равномерной сходимости почти наверное [1] $P\left(\sup_x |f_n(x) - f(x)| \xrightarrow[n\to\infty]{} 0\right) = 1$: достаточно, чтобы существовало такое n, что $m^S(L) \le L^n + 1$.
В нашем случае $n = 1$, и факт доказан.

Задача 2

А) Подвыборка m различных целых чисел, индуцированная данным классом решающих правил, однозначно определяется как пара (i, j), $0 \le i < j \le m$ - индексы вставки чисел a, b в возрастающую последовательность элементов выборки, если выделяемая подвыборка не пустая. Поскольку количество таких различных пар равно C_{m+1}^2, и учитывая случай пустой выделяемой подвыборки, получаем

$$m^S(L) = C_{L+1}^2 + 1$$

В) Этот класс подвыборок содержит предыдущий, и помимо этого порождает подвыборки, где a, b, c, d соответствуют индексам (i, j, k, l), $0 \le i < j < k < l \le m$. Получаем

$$m^S(L) = C_{L+1}^2 + 1 + C_{L+1}^4$$

C) Поскольку пересечение двух отрезков тоже отрезок, то здесь такой же ответ, как в пункте A:

$$m^S(L) = C_{L+1}^2 + 1$$

Задача 3

$vc_n = \max\{L | m_n^S(L) = 2^L\}$−?
Где $x \in \mathbb{R}^n$, S - множество, порожденное разбиениями всевозможных гиперплоскостей.

Для фиксированных точек x_i нас интересуют решения неравенств $\langle w, x_i \rangle + b \lesseqgtr 0$ относительно w, b. Но в таком виде задача некорректно поставлена. Регуляризуем ее следующим образом:
$c_i(\langle w, x_i \rangle + b) \geq 1; ||w|| \to min$, где $c_i = \pm 1$ - класс объекта.
Пусть $x_0 = 0, x_i = e_i; \hat{x} = (x, 1), \hat{w} = (w, b)$. Тогда $\hat{X}\hat{w} = y$ имеет решение для любого y, поскольку расширенная матрица $\hat{X}$ обратима, значит $m_n^S(n+1) = 2^{n+1}$.

А поскольку размерность пространства y не может быть больше размерности пространства переменных, то $m_n^S(n+2) < 2^{n+2}$, и $vc_n = n + 1$.

[1] В.Н.В а п н и к , А . Я . Ч е р в о н е н к и с , О равномерной сходимости частот появления событий к их вероятностям, Д А Н СССР, 181, 4 (1968), 781.