Задание 3 (I курс, весенний семестр 2010г.)

Задача 1 (обязательная для всех).

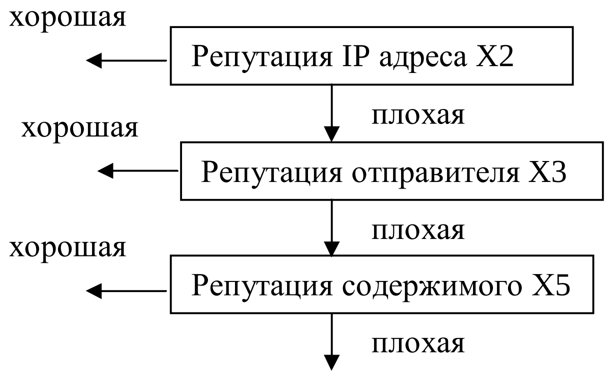
Репутационный метод выявления спама основан на признаках

- X1 Репутация домена
- X2 Репутация адреса IP
- X3 Репутация отправителя
- X4 Репутация получателя
- X5 Репутация содержимого письма

Будем считать перечисленные признаки двоичными (0-плохая репутация; 1 – хорошая репутация) и обозначим $P(S \mid Xi = 0)$ вероятность того, что сообщение, для которого признак Xi равен нулю, является спамом, а через $P(S \mid Xi = 1)$ вероятность того, что сообщение, для которого признак Xi равен единице, является спамом.
Допустим, что условная вероятность того, что сообщение с набором признаков X1, X2, X3, X4, X5 является спамом, задаётся соотношением

$$P(S \mid X1, X2, X3, X4, X5) = P(S \mid X1)P(S \mid X2)P(S \mid X3)P(S \mid X4)P(S \mid X5)$$

- записать выражение для вероятности того, что пришедшее письмо является спамом
- записать выражение для вероятности того, что пришедшее письмо является спамом, если в сети равновероятно присутствуют адреса IP с плохой и с хорошей репутацией. Все отправители с плохой репутацией принадлежат только адресам IP с плохой репутацией, но среди отправителей, принадлежащих адресам IP с плохой репутацией, равновероятно присутствуют и отправители с хорошей репутацией. Все письма с плохой репутацией отправляются только отправителями с плохой репутацией, но содержания писем, полученных от отправителя с плохой репутацией, равновероятно могут иметь как плохую, так и хорошую репутацию (см. рисунок).



Задача 2

Пусть коэффициенты разложения $f(x) = a_0 + \sum_{k=1}^{25} \left(a_k \sin(2pkx) + b_k \cos(2pkx)\right)$ являются независимыми нормально распределёнными величинами со средними значениями 0 и дисперсиями $s_k^2 = \frac{1}{1+0.12k}$, k=0,…,25. Сгенерировать значения одной реализации функции $f(x)$ на равномерной сетке с узлами $x_i = -1 + \frac{i}{25}$, i=1,…,50. Вычислить вектор u с элементами $u_i = f(t_i) + x_i$, для $t_i = -1 + \frac{2i}{25}$ i=1,…,25, где x_i - случайные нормально распределённые величины, имеющие средние значения 0 и дисперсию $s_x^2 = 0.09$.

Вычислить квадратичное уклонение $F(s) = \sum_{i=1}^{25} \left(f(z_i) - f_i^s\right)^2$, где $f(z_i)$ - значения сгенерированной ранее реализации функции $f(x)$ в точках $z_i = -1 + \frac{2i-1}{25}$, i=1,…,25, f_i^s - апостериорное среднее значение случайной величины $f(z_i)$, при наблюдённом векторе u, s^2 - дисперсия возмущения x_i, которая считается неизвестной.

Вычислить величину $F(s)$ при различных значениях s, включая $s = 0.3$.

- Построить график $F(s)$ в зависимости от s. Где достигается минимум построенной зависимости?
- Повторить вычисления для различных реализаций коэффициентов разложения.
- Сделать выводы.

Задача 3

Корреляционная функция стационарного трёхмерного поля (гауссова с нулевым средним) равна $R_e(x,y) = (1-e)e^{-0.3\|x-y\|^2} + ed\left(\|x-y\|^2\right)$.

Сгенерировать реализацию значений поля u_1 в узлах

(-1,-1,0); (-1,0,0); (-1,1,0); (0,-1,0); (0,0,0); (0,1,0); (1,-1,0); (1,0,0); (1,1,0);

(-1,-1,1); (-1,0,1); (-1,1,1); (0,-1,1); (0,0,1); (0,1,1); (1,-1,1); (1,0,1); (1,1,1)

при $e = 0.1$. Методом кригинга **по реализации поля в узлах** (-1,-1,0); (-1,0,0); (-1,1,0); (0,-1,0); (0,0,0); (0,1,0); (1,-1,0); (1,0,0); (1,1,0); спрогнозировать значения поля u_e в узлах (-1,-1,1); (-1,0,1); (-1,1,1); (0,-1,1); (0,0,1); (0,1,1); (1,-1,1); (1,0,1); (1,1,1) для разных значений e, включая $e = 0.1$. Для каждого значения e вычислить величину $F(e) = \|u_2 - u_e\|^2$, где u_2 значения **сгенерированной ранее реализации поля** u_1 в узлах (-1,-1,1); (-1,0,1); (-1,1,1); (0,-1,1); (0,0,1); (0,1,1); (1,-1,1); (1,0,1); (1,1,1).

- Построить график $F(e)$ в зависимости от e. Где достигается минимум построенной зависимости?
- Повторить вычисления для различных реализаций поля.
- Сделать выводы.

Задача 4

На вход линейной системы с передаточной функцией $K(t,t) = \exp(-0.3(t-t))$ поступает белый шум (случайный процесс с некоррелированными значениями). Записать формулу для прогнозирования методом кригинга значений выходного сигнала в моменты времени $t_i^* = i + 0.5$ по значениям выходного сигнала в моменты времени $t_i = i$, i=1,…,10.

Указание. Выход линейной системы связан с входом соотношением

$$U_{вых}(t) = \int_0^t K(t,t)U_{вход}(t)dt$$