Задание 4 (I курс, весенний семестр 2010г.)

Задача 1.

В вершинах равностороннего треугольника проводятся измерения, по которым методом кригинга прогнозируется значение поля в центре треугольника. Пусть поле стационарно, то есть дисперсия его значения в любой точке постоянна, а корреляционная функция зависит только от расстояния между двумя точками. Вычислить величину прогнозируемого методом кригинга значения поля в центре треугольника и дисперсию этого значения.

Задача 2.

Записать ковариационную матрицу для оценки методом гребневой регрессии по данным, искажённым независимой помехой с нулевым средним значением и дисперсией, равной 2.

Задача 3.

Записать математическое ожидание квадрата уклонения от истинной величины для оценки методом гребневой регрессии по данным, искажённым независимой помехой с нулевым средним значением и дисперсией, равной 2.

Задача 4.

Случайные точки из двух классов располагаются на n-мерной сфере. Точки класса 0 с вероятностью 1 располагаются на полусфере x1>0, точки класса 1 с вероятностью 1 располагаются на полусфере x1≤0. Точки классифицируются на два класса с помощью произвольных выпуклых множеств (точки, принадлежащие множеству, относятся к классу 0, остальные к классу 1).

Можно ли гарантировать, что при достаточно большом числе обучающих примеров будет достигаться минимум среднего числа ошибок классификации на всех множествах, минимизирующих число ошибок на обучающей выборке? **Ответ обосновать.**

Задача 5.

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 $x_i = -1 + \frac{i}{50}$, i=1,…,100. Вычислить вектор u с элементами $u_i = f(t_i) + x_i$, для $t_i = -1 + \frac{i}{25}$ i=1,…,50, где x_i - случайные нормально распределённые величины, имеющие средние значения 0 и дисперсию $s_x^2 = 0.09$.

Вычислить квадратичное уклонение $F(s) = \sum_{i=1}^{50} \left(f(z_i) - f_i^s\right)^2$, где $f(z_i)$ - значения сгенерированной ранее реализации функции $f(x)$ в точках $z_i = -1 + \frac{2i-1}{50}$, i=1,…,50, 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.
- Описать процедуру определения величины s - среднеквадратичного уклонения шума x_i, используя комбинированный подход: максимум правдоподобия – Байес (материал лекции от 19.02.2010 г.). Найти оценку s, максимизирующую правдоподобие.
- Сравнить значения $F(s)$ для $s = 0.3$ и при оценке s, найденной, используя комбинированный подход.
- Повторить вычисления для различных реализаций коэффициентов разложения.
- Сделать выводы.

Задача 6.

Регрессионная зависимость приближается с помощью только одного члена ряда Фурье путём минимизации по a и по k эмпирического риска $J_e(a,k) = \frac{1}{l}\sum_{i=1}^{l}(y_i - af_k(x_i))^2$, где $(x_1, y_1, ..., x_l, y_l)$ - независимая выборка значений пары (x, y), имеющей распределение $P(x, y)$, $f_k(x)$ - k – ый член ряда Фурье.

Можно ли утверждать, что при достаточно большом числе l минимум эмпирического риска $J_e(a^*, k^*) = \min_{a,k} J_e(a,k)$ сколь угодно близок к величине среднего риска $J_m(a^*, k^*) = \int (y - a^* f_{k^*}(x))^2 dP(x, y)$? **Ответ обосновать.**