Задание 5

Выполнил Циглер А.С.

Задача 1

$x : p(x) \sim \exp(-\frac{(x-\mu)^T\Lambda(x-\mu)}{2})$

$y|x : p(y|x) \sim \exp(-\frac{(y-Ax-b)^T L(y-Ax-b)}{2})$

Понятно, что по $p(x), p(y|x)$ восстанавливается однозначно совместное распределение и распределения $p(y), p(x|y)$.

Заметим, что если выбросить значение $x \sim \mathcal{N}(\mu, \Lambda^{-1})$, а затем выбросить значение $\varepsilon \sim \mathcal{N}(0, L^{-1})$, то для случайной величины $y' = Ax + b + \varepsilon$ будет выполнено $p(y'|x) = p(y|x)$. То есть можно считать $y = y'$, от этого ответ в задаче не изменится.

x, ε - независимые случайные векторы, ε - центрированный. $Ax + b \sim \mathcal{N}(A\mu + b, A\Lambda^{-1}A^T)$

$Ax + b + \varepsilon$ - очевидно, гауссовский вектор с ожиданием $A\mu + b$, надо лишь найти ковариацию:

$E(Ax - A\mu + \varepsilon)(Ax - A\mu + \varepsilon)^T = E(Ax - A\mu)(Ax - A\mu)^T + 2E(Ax - A\mu)\varepsilon^T + E\varepsilon\varepsilon^T = E(Ax - A\mu)(Ax - A\mu)^T + E\varepsilon\varepsilon^T = A\Lambda^{-1}A^T + L^{-1}$, итого $y \sim \mathcal{N}(A\mu + b, A\Lambda^{-1}A^T + L^{-1})$

Теперь мы хотим плотность x при условии известного y. То есть мы знаем, что сумма независимых векторов $Ax + b$ и ε равна y, то есть $\varepsilon = y - b - Ax$. Для данной пары (x, ε) плотность в этой точке пропорциональна $\exp(-\frac{(x-\mu)^T\Lambda(x-\mu)}{2})\exp(-\frac{(y-Ax-b)^T L(y-Ax-b)}{2}) \sim \exp(-\frac{x^T(\Lambda+A^TLA)x - 2x^T(\Lambda\mu + A^T(y-b))}{2})$. Стало быть, обозначив $\Sigma = (\Lambda + A^TLA)^{-1}$ получаем $p(x|y) \sim \mathcal{N}(\Sigma(\Lambda\mu + A^T(y - b)), \Sigma)$ - прям как просили.

Задача 2

Верхними индексами будем обозначать координаты вектора.

Возьмем очень далёкие точки x, x'. Тогда $k(x, x) = k(x', x') = 1$, $k(x, x') = k(x', x) = exp(-\theta(x - x')^2)$ - большое положительное число. Но тогда $k(x, x)k(x', x') - k(x, x')^2 < 0$ - нет положительной определенности. То есть при таком θ k не будет ковариационной функцией.

Задача 3

Как я понял, имеется ввиду, что мы считаем $y = f(x) + \varepsilon$, где $f(x)$ - реализация гауссовского процесса с нулевым средним и ковариационной функцией $k(x, x') = x^Tx'$, а ε - белый гауссовский шум. Для такой модели $y(x_{m+1})$ ищется по методу максимального правдоподобия.

Давайте поймем, что это за гауссовский процесс: мы знаем, что для задания гауссовского процесса достаточно задать среднее и ковариацию. Но рассмотрим такой процесс: $f'(x) = \sum\limits_1^n x^i\xi_i$, где ξ_i - независимые стандартные нормальные случайные величины (одномерные). Такой процесс будет гауссовским. Действительно, рассмотрим сечение $(f'(x_1), \ldots, f'(x_m)) = [x_1, x_2, \ldots, x_m]^T\xi$, где ξ - вектор тех самых стандартных нормальных величин. То есть каждое сечение - это линейное преобразование гауссовского вектора, каждое сечение - гауссовский вектор. Но давайте теперь посмотрим на ковариацию: $Ef(x_1)f(x_2) = E(\sum\limits_i x_1^i\xi_i)(\sum\limits_j x_1^j\xi_j) = \sum x_1^i x_2^i E\xi_i^2 = x_1^T x_2$ - прям как у нас.

Значит, если мы будем считать $f(x) = \sum\limits_1^n x^i\xi_i$, то модель не изменится. То есть мы считаем $y = \sum x^i\xi^i + \varepsilon$ - а это модель регрессии $y = \theta^T x + \varepsilon$, но еще на параметр θ есть априорное нормальное распределение.

То есть в итоге мы ищем вектор θ минимизируя $\sigma^{-2}\sum(y_i - \theta^T x_i)^2 + \theta^T\theta$

Задача 4

Что у нас вообще происходит: ясно, что от всего гауссовского процесса можно оставить только проекцию на точки $\{x_i\}$. Получается гауссовский вектор с ковариационной матрицей $K : K[i, j] = k(x_i, x_j)$.

Дальше можно забыть про иксы. То есть у нас есть вектор y гауссовский с данной ковариационной матрицей. Мы по очереди берем каждую координату и сравниваем с ее оценкой максимального правдоподобия и смотрим, что получилось. Плотность монотонно зависит от $y^T K^{-1} y$. Обозначим за e_j j-ый координатный вектор. Нас интересует минимум $(y + te_j)^T K^{-1}(y + te_j)$. Ну, это парабола. $2te_j^T K^{-1} e_j = 2e_j^T K^{-1} y$, $(\widehat{y}_{-j} - y_j)^2 = t^2 = (\frac{e_j^T K^{-1} y}{e_j^T K^{-1} e_j})^2$ - то есть надо взять от матрицы K^{-1} j-ую сторку, поделить это дело на j-ый диагональный элемент матрицы K^{-1} и умножить на вектор y. Нас интересует сумма квадратов этого по всем j. Обозначим за Λ диагональную матрицу, у которой диагональ совпадает с матрицей K^{-1}.

В итоге получили $\sum(\widehat{y}_{-j} - y_j)^2 = \|\Lambda^{-1} K^{-1} y\|^2$