2.11 Ликбез по линалу. Многомерные распределения. Регрессия. Катарсис.

Пререквизиты

Вектор - столбец. n - количество объектов, d - количество признаков (размерность пространства)

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$$

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_n^T \end{pmatrix} = \begin{pmatrix} x_{1,1} & \dots & x_{1,d} \\ \vdots & & \\ x_{n,1} & \dots & x_{n,d} \end{pmatrix}$$

Матожидание вектора(для матрицы - аналогично):

$$\mathbb{E}\boldsymbol{x} = \begin{pmatrix} \mathbb{E}x_1 \\ \vdots \\ \mathbb{E}x_n \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix}$$

Ковариация двух векторов:

$$\mathrm{Cov}(\boldsymbol{x}, \boldsymbol{y}) = \begin{pmatrix} \mathrm{Cov}(x_1, y_1) & \dots & \mathrm{Cov}(x_1, y_d) \\ \vdots & \dots & \vdots \\ \mathrm{Cov}(x_d, y_1) & \dots & \mathrm{Cov}(x_d, y_d) \end{pmatrix}$$

Дисперсия - ковариация одного и того же вектора

$$\mathbb{V}\boldsymbol{x} = \begin{pmatrix} \mathrm{Cov}(x_1, x_1) & \dots & \mathrm{Cov}(x_1, x_d) \\ \vdots & \dots & \vdots \\ \mathrm{Cov}(x_d, x_1) & \dots & \mathrm{Cov}(x_d, x_d) \end{pmatrix} = \mathrm{Cov}(\boldsymbol{x}, \boldsymbol{x})$$

Ну и понятно как эти вещи расширять на еще большие пространства.

Многомерные распределения Плотность многомерного равномерного распределения.

Пусть есть некое борелевское (ограниченное непрерывное) множество векторов S. $\lambda(S)$ - объем множества (конечная мера Лебега). Тогда плотность определяется:

$$\rho(\boldsymbol{x}) = \begin{cases} \frac{1}{\lambda(S)} & \boldsymbol{x} \in S \\ 0 & overwise \end{cases}$$

Плотность многомерного нормального распределения: Пусть у нас пространство расмерности d. Определим плотность нормального распределения с матожиданием μ и матрица ковариаций Σ

$$\rho(\boldsymbol{x}) = \frac{1}{\sqrt{\det \Sigma}(\sqrt{2\pi})^n} \exp\Big(\frac{1}{2}(\boldsymbol{x}-\mu)^T\Sigma^{-1}(\boldsymbol{x}-\mu)\Big)$$

Осталось осознать, что многомерные распределения, это частный случай совместных распределений, которые мы уже изучали в прошлом семестре:)

Производные по матрице

Напомним что такое градиент и гессиан функции $F(x_1, x_2, \ldots, x_d)$

$$\nabla_{\boldsymbol{x}} F = \Big(\frac{\partial F}{\partial x_1}, \ldots, \frac{\partial F}{\partial x_d}\Big)$$

$$\nabla^2 F = \begin{pmatrix} \frac{\partial^2 F}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 F}{\partial x_1 \partial x_d} \\ \vdots & \cdots & \vdots \\ \frac{\partial^2 F}{\partial x_d \partial x_1} & \cdots & \frac{\partial^2 F}{\partial x_d \partial x_d} \end{pmatrix}$$

Ну наверно никто не удивится тому, что

$$\nabla_{\boldsymbol{x}} F = \frac{\partial F}{\partial \boldsymbol{x}}$$

2.11 Ликбез по линалу. Многомерные распределения. Регрессия. Катарсис.

Практика

1. Мы хотим восстановить некоторую функцию f(x) по данным аргументам x_i и значениям функции y_i. Используем метод линейной регрессии:

$$\boldsymbol{y} = \beta_1 \boldsymbol{x} + \beta_0 + \boldsymbol{\varepsilon}$$

   $\boldsymbol{\varepsilon}$ - некий шум, который распределен нормально. Задача регрессии:

$$\sum_{i=1}^{n} \varepsilon_i^2 \to \min_{\beta_0, \beta_1}$$

   Решите задачу регрессии: найдите параметры β_1 и β_0 и покажите их состоятельность.

2. Теперь у нас многомерная функция $f(x_1, x_2, \ldots, x_d)$. Матрица аргументов - $\boldsymbol{X}$ и ее значения $\boldsymbol{y}$. Решим задачу линейной регрессии для многомерной функции:

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon}$$
$$\sum_{i=1}^{n} \varepsilon_i^2 \to \min_{\beta}$$

3. (Ridge Regression) Теперь поборемся с хреновой обусловленностью матрицы $X^T X$.

$$\hat{\beta} = \arg\min_{\beta} ||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||^2 + \lambda \boldsymbol{\beta}^T \boldsymbol{\beta}$$

4. Докажем, что задача регрессии эквивалентна задаче нахождения параметров методом максимального правдоподобия.

2.11 Ликбез по линалу. Многомерные распределения. Регрессия. Катарсис.

Домашка

1. (4) Понятное дело, считать обратные матрицы - это полная жесть. Поэтому вам надо записать формулы для нахождения $\boldsymbol{\beta}$ с помощью градиентного спуска для обычной и ridge регрессии. (Формулу градиентного спуска ищите в википедии яустал техать)
2. Доказать градиент и гессиан функций:
   (a) (2) (Квадратичная функция) $f(\boldsymbol{x}) = \frac{1}{2}(Ax)^T x - b^T x + c$
   (b) (3) Куб нормы $\frac{1}{3}||x||_2^3$
3. Докажите, что (1)суммарная мера по всему пространству многомерного нормального распределения равна 1, (1) - матожидание, (2)Σ - матрица ковариаций. Да тут будет интегрирование по вектору (а кажется просто интегрирования по нескольким переменным). Достаточно двумерного распределения, если докажете для общего баллы умножатся на 2.
4. (1)А что будет, если добавить шум не только к $\boldsymbol{y}$, но и к $\boldsymbol{x}$? Пусть нам известны $\boldsymbol{y}$ и $\boldsymbol{Z} = \boldsymbol{x} + \delta$. Выпишете решение обычной задачи регрессии в этом случае. (3)Будет ли оценка состоятельной?(Спойлер нет, докажите)