A3 HW theory

Фокин Алексей, 922 группа

23 октября 2021 г.

Задача 1

До измерений мы ничего не знаем, поэтому априорную вероятность будем считать постоянной для простоты:

$$p(\lambda) = \frac{1}{\Lambda}$$

Согласно условию,

$$P_\lambda(n) \equiv P(n \mid \lambda) = \frac{\lambda^n}{n!}e^{-\lambda}$$

Из теоремы Байеса получаем

$$P(\lambda \mid m) = \frac{P(m \mid \lambda)P(\lambda)}{\sum\limits_{\lambda'} P(m \mid \lambda')P(\lambda')} = \frac{p(\lambda)P(m \mid \lambda)}{\int_0^\infty p(\lambda')P(m \mid \lambda')d\lambda'} = \frac{P(m \mid \lambda)}{\int_0^\infty \frac{\lambda'^m}{m!}e^{-\lambda'}d\lambda'} = P(m \mid \lambda) = \frac{\lambda^m}{m!}e^{-\lambda}$$

Замечательно. Поскольку это все наши данные, будем считать $p(\lambda) = P(\lambda \mid m)$ новой априорной вероятностью $p(\lambda)$

Это наблюдение мы вольны применить к формуле Байеса, тогда

$$P\left(\lambda \mid m'\right) = \frac{\lambda^m}{m!}e^{-\lambda}\frac{\frac{\lambda^{m'}}{m!}e^{-\lambda}}{\int_0^\infty \frac{\lambda'^m}{m!}e^{-\lambda'} \cdot \frac{\lambda'^{m'}}{m'!}e^{-\lambda'}d\lambda'} = \frac{(2\lambda)^{m+m'}}{m'!m!}e^{-2\lambda} \cdot \frac{2m!m'!}{(m+m')!} = 2\frac{(2\lambda)^{m+m'}}{(m+m')!}e^{-2\lambda}$$

Задача 2

Переведём условие на язык Байеса:

$$\begin{aligned} P(\text{болен}) &= 10^{-5} \\ P(\text{положительный тест} \mid \text{болен}) &= 0.01 \\ P(\text{отрицательный тест} \mid \text{здоров}) &= 0.01 \end{aligned}$$

Вероятность того, что Петя болен:

$$P(\text{болен} \mid \text{положительный тест}) = \frac{P(\text{положительный тест} \mid \text{болен})P(\text{болен})}{\sum_{B=\text{болен, здоров}} P(\text{положительный тест} \mid B)P(B)}$$

Будем считать $P(\text{положительный тест} \mid \text{здоров}) \approx 0.99$, тогда

$$P(\text{болен} \mid \text{положительный тест}) = \frac{0.99 * 10^{-5}}{0.99 * 10^{-5} + 0.01 * 0.9999} \approx 10^{-3}$$

Задача 5

Минимизируем $L(w) = \|Xw - y\|_2$.
Введём функцию $g = \|w\|_1 - C$.

Задача Лассо Тибширани формулируется как

$$L \to min, \quad g \le 0$$

Условия Каруша-Куна-Таккера работают для функций $L(\bar{x})$, т.ч. $x_\alpha \ge 0$, и мы действительно можем переформулировать нашу задачу так:

$$w_a = u_a - v_a \equiv \frac{|w_a| + w_a}{2} - \frac{|w_a| - w_a}{2},$$

тогда $L(w) = L(u, v), \quad u_a, v_a \ge 0$

Согласно условиям ККТ,

$$L \to min, \quad g \le 0 \quad \Rightarrow \quad \exists \lambda \ge 0 : \min_w L(w) = L(\hat{w}) + \lambda g(\hat{w}), \quad \lambda g(\hat{w}) = 0, \quad \hat{w} \in \text{argmin} L$$

$$\exists \lambda > 0 : \min_w L(w) = L(\hat{w}) + \lambda g(\hat{w}), \quad \lambda g(\hat{w}) = 0, \quad \Rightarrow \quad L \to min, g \le 0$$

То есть там, где выполняются эти условия, $L(w) + \lambda g(w) = L(w)$, тогда Тибширани $\Rightarrow_{KKT}$ L1.

L1-регуляризация же есть минимизация $L_1(w) = L(w) + \lambda g(w)$ без доп. условий, и тогда в стац. точке выполняется $\partial_\lambda L_1 = g(\hat{w}) = 0 \quad \Rightarrow_{KKT}$ Тибширани.

Задача 6

Попытаемся "облегчить"обозначения, используя $\mathbb{E}$ и $\mathbb{E}'$ в качестве $\mathbb{E}_{x,y}$ и $\mathbb{E}_{X_l, y_l}$, понимая, где что.

Простой шаг:

$$\mathbb{E}[(y - \hat{y})^2] = \mathbb{E}[(y - \mathbb{E}[y|x] + \mathbb{E}[(y|x)] - \hat{y})^2] =$$
$$\mathbb{E}[(y - \mathbb{E}[y|x])^2] + \mathbb{E}[(\mathbb{E}[y|x] - \hat{y})^2] + 2\mathbb{E}[(y - \mathbb{E}[y|x])(\mathbb{E}[y|x] - \hat{y})];$$
$$\mathbb{E}[(y - \mathbb{E}[y|x]) \cdot (\mathbb{E}[y|x] - \hat{y})] = \mathbb{E}[(y - \mathbb{E}[y|x])] \cdot \mathbb{E}[(\mathbb{E}[y|x] - \hat{y})],$$

т. к. y и $\hat{y}$.

Теперь получим слагаемое с вариацией выборки:

$$\mathbb{E}[(\mathbb{E}[y|x] - \hat{y})^2] = \mathbb{E}[(\mathbb{E}[y|x] - \mathbb{E}'[\hat{y}] + \mathbb{E}'[\hat{y}] - \hat{y})^2] =$$
$$\mathbb{E}[(\mathbb{E}'[\hat{y}] - \mathbb{E}[y|x])^2] + \mathbb{E}[(\mathbb{E}'[\hat{y}] - \hat{y})^2] + 2\mathbb{E}[(\mathbb{E}'[\hat{y}] - \mathbb{E}[y|x])(\mathbb{E}'[\hat{y}] - \hat{y})] =$$
$$\mathbb{E}[(\mathbb{E}'[\hat{y}] - \mathbb{E}[y|x])^2] + \mathbb{E}[(\mathbb{E}'[\hat{y}] - \hat{y})^2] + 2(\mathbb{E}'[\hat{y}] - \mathbb{E}[y|x])(\mathbb{E}\mathbb{E}'[\hat{y}] - \mathbb{E}[\hat{y}])$$

Посмотрим на среднее по выборке от перекрёстных слагаемых:

$$\mathbb{E}'[\mathbb{E}[(y - \mathbb{E}[y|x])] \cdot \mathbb{E}[(\mathbb{E}[y|x] - \hat{y})] + (\mathbb{E}'[\hat{y}] - \mathbb{E}[y|x])(\mathbb{E}\mathbb{E}'[\hat{y}] - \mathbb{E}[\hat{y}])] =$$
$$(\mathbb{E}'\mathbb{E}[y] - \underbrace{\mathbb{E}'\mathbb{E}[y|x]}_{\mathbb{E}'\mathbb{E}[y]}) \cdot (\mathbb{E}'\mathbb{E}[y|x] - \mathbb{E}'\mathbb{E}[\hat{y}]) + (\mathbb{E}'[\hat{y}] - \mathbb{E}'\mathbb{E}[y|x])(\mathbb{E}\mathbb{E}'[\hat{y}] - \mathbb{E}'\mathbb{E}[\hat{y}]) = 0 + 0 = 0$$

Итого,

$$\mathbb{E}'\mathbb{E}[(y-\hat{y})^2] = \mathbb{E}[(y-\mathbb{E}[y|x])^2] + \mathbb{E}[(\mathbb{E}'[\hat{y}]-\mathbb{E}[y|x])^2] + \mathbb{E}[(\mathbb{E}'[\hat{y}]-\hat{y})^2],$$

QED.