

Задачи к лекции и семинару “Регуляризация”

1

Количество срабатываний счетчика Гейгера за минуту n подчиняется распределению Пуассона:

$$P_{\lambda}(n) = \frac{\lambda^n}{n!} e^{-\lambda}.$$

1. В ходе эксперимента счетчик Гейгера сработал за минуту m раз. С помощью теоремы Байеса определите апостериорное распределение на λ .
2. Эксперимент повторили еще раз, в этот раз счетчик Гейгера сработал за минуту m' раз. Как изменилось апостериорное распределение на λ ?

2

Ультразвуковой тест от коронавируса ошибается в 1% случаев (как в одну, так и в другую сторону). В данный момент в популяции доля заболевших 10^{-5} . Петя получил положительный тест на коронавирус. С какой вероятностью он действительно болеет коронавирусом?

3

На семинаре обсуждалось решение задачи регрессии с $L1$ -регуляризацией с помощью метода градиентного спуска. С помощью K-Fold кроссвалидации¹ ($K = 3$) осуществите для этого метода подбор параметров: коэффициент перед регуляризатором и параметр градиентного спуска (learning rate). В качестве данных возьмите значения какой-нибудь неполиномиальной функции на равномерной или случайной сетке (на выбор семинариста) с добавленным гауссовым шумом. Насколько стабильно по отношению к запуску работает градиентный спуск?

4

Для стандартного набора данных для задачи регрессии (см. например `load_diabetes` из `sklearn.datasets`) продемонстрируйте, как веса обращаются в ноль по мере увеличения коэффициента $L1$ -регуляризации. Разрешается использовать библиотечную реализацию регрессии.

5

Покажите, что задача минимизации квадратичной функции потерь с дополнительным ограничением (лассо Тибширани):

$$\mathcal{L} = \|Xw - y\|^2 \rightarrow \min_w, \quad \sum_{\alpha} |w_{\alpha}| < C$$

эквивалентна $L1$ -регуляризации. *Указание:* можно воспользоваться условиями Каруша — Куна — Таккера (обобщение метода Лагранжа). ([link](#)).

¹Способ кроссвалидации, при котором данные разбиваются на K частей и совершается K запусков обучения: по очереди каждая из частей объявляется тестовой, а объединение оставшихся $K - 1$ частей используется для обучения. Такой способ обсуждался на первой лекции.

6*. Bias-Variance decomposition

Воспользуемся вероятностной моделью данных, в которой предполагается, что каждый элемент выборки независимо от других поступает из распределения $p(x, y)$. Тогда вероятность получить какой-то конкретный набор данных $(X_l, y_l) = (x_1, \dots, x_l; y_1, \dots, y_l)$ в обучающей выборке равна $p(X_l, y_l) = \prod_{i=1}^l p(x_i, y_i)$. В дальнейшем будем обозначать как (x, y) элемент тестовой выборки, который не входит в (X_l, y_l) .

В выбранной модели $\tilde{y} = g_\theta(x)$ параметры θ определяются с помощью фиттирования по обучающей выборке: $\theta = \theta(X_l, y_l)$, поэтому \tilde{y} зависит от x , X_l и y_l . Тогда формальное выражение для функции потерь (соответствующее пределу бесконечной большой тестовой выборки) можно записать как

$$L = \mathbb{E}_{X_l, y_l} \left[\mathbb{E}_{x, y} (y - \tilde{y})^2 \right].$$

В этом выражении квадратичная функция потерь усредняется по элементу тестовой выборки (x, y) и по обучающей выборке (X_l, y_l) .

Покажите, что справедливо разложение этой величины на шум, смещение и разброс:

$$L = \underbrace{\mathbb{E}_{x, y} (y - \mathbb{E}(y|x))^2}_{\text{noise}} + \underbrace{\mathbb{E}_{x, y} (\mathbb{E}_{X_l, y_l}(\tilde{y}) - \mathbb{E}(y|x))^2}_{\text{bias}} + \underbrace{\mathbb{E}_{x, y} \left[\mathbb{E}_{X_l, y_l} (\tilde{y} - \mathbb{E}_{X_l, y_l} \tilde{y})^2 \right]}_{\text{variance}}$$

Указание: сначала покажите, что

$$\mathbb{E}_{x, y} (y - \tilde{y})^2 = \mathbb{E}_{x, y} (y - \mathbb{E}(y|x))^2 + \mathbb{E}_{x, y} (\mathbb{E}(y|x) - \tilde{y})^2$$