Машинное обучение, ФКН ВШЭ
Теоретическое домашнее задание №1
Линейные модели

Задача 1. Скоро первая самостоятельная работа. Чтобы подготовиться к ней, ФКН ест конфеты и решает задачи. Число решённых задач y зависит от числа съеденных конфет x. Если студент не съел ни одной конфеты, то он не хочет решать задачи. Поэтому для описания зависимости числа решённых задач от числа съеденных конфет используется линейная модель с одним признаком без константы $y_i = w \cdot x_i$. В аналитическом виде найдите оценки параметра w, минимизируя следующие функции потерь:

1. Линейная регрессия без штрафа: $Q(w) = \frac{1}{\ell}\sum_{i=1}^{\ell}(y_i - wx_i)^2$;
2. Ridge-регрессия: $Q(w) = \frac{1}{\ell}\sum_{i=1}^{\ell}(y_i - wx_i)^2 + \lambda w^2$;
3. LASSO-регрессия: $Q(w) = \frac{1}{\ell}\sum_{i=1}^{\ell}(y_i - wx_i)^2 + \lambda|w|$;
4. Пусть решения этих задач равны $\hat{w}, \hat{w}_R$ и $\hat{w}_L$ соответственно. Найдите пределы

$$\lim_{\lambda\to 0}\hat{w}_R, \quad \lim_{\lambda\to\infty}\hat{w}_R, \quad \lim_{\lambda\to 0}\hat{w}_L, \quad \lim_{\lambda\to\infty}\hat{w}_L.$$

5. Как можно проинтерпретировать гиперпараметр λ?

Hint: в случае Lasso-регрессии придётся повозиться с модулем. Обратите внимание на то, что $Q(w)$ парабола, это поможет корректно найти аналитическое решение. Подумайте, с чем возникнут проблемы, если у нас будет не один параметр, а сотня.

Задача 2. Вася измерил вес трёх покемонов, $y_1 = 6$, $y_2 = 6$, $y_3 = 10$. Вася хочет спрогнозировать вес следующего покемона с помощью константной модели $y_i = w$. Для оценки параметра w Вася использует целевую функцию

$$\frac{1}{\ell}\sum_{i=1}^{\ell}(y_i - w)^2 + \lambda w^2.$$

1. Найдите оптимальное w при произвольном λ.

2. Подберите оптимальное λ с помощью кросс-валидации leave one out («выкинь одного»). На первом шаге мы оцениваем модель на всей выборке без первого наблюдения, а на первом тестируем её. На втором шаге мы оцениваем модель на всей выборке без второго наблюдения, а на втором тестируем её. И так далее ℓ раз. Чтобы найти λ_{CV} мы минимизируем среднюю ошибку, допущенную на тестовых выборках.

3. Найдите оптимальное значение w при λ_{CV}, подобранном на предыдущем шаге.

4. Выведите формулу для λ_{CV} при произвольном количестве наблюдений.

Задача 3. Убедитесь, что вы знаете ответы на следующие вопросы:

- Что такое гиперпараметр модели и чем он отличается от параметра модели?
- Почему коэффициент регуляризации нельзя подбирать по обучающей выборке? Как подобрать оптимальное значение для коэффициента регуляризации?
- Почему накладывать регуляризатор на свободный коэффициент w_0 может быть плохой идеей?
- Что такое кросс-валидация, чем она лучше использования отложенной выборки?
- Почему категориальные признаки нельзя закодировать натуральными числами? Что такое one-hot encoding?
- Для чего нужно масштабировать матрицу объекты-признаки перед обучением моделей машинного обучения?
- Почему L_1-регуляризация производит отбор признаков?
- Почему MSE чувствительно к выбросам?