

ML. Первое занятие

План занятия

1. Постановка задачи машинного обучения
2. Основные понятия
3. Типы признаков
4. Типы задач машинного обучения
5. Задача регрессии
6. Линейные модели
7. Линейная регрессия
8. Подготовка признаков
9. Измерение ошибки в задачах регрессии
10. Обобщающая способность модели
11. Борьба с переобучением
12. Регуляризация
13. Обучение линейной модели
14. Градиентный спуск
15. Оптимизации градиентного спуска

Постановка задачи машинного обучения

Постановка задачи машинного обучения

Задача: определение спама по входящим звонкам.

Постановка задачи машинного обучения

Задача: определение спама по входящим звонкам.

Семейство функций(алгоритмы), которые на вход получают аудио звонка и выдают результат: спам или не спам.

Постановка задачи машинного обучения

Задача: определение спама по входящим звонкам.

Семейство функций(алгоритмы), которые на вход получают аудио звонка и выдают результат: спам или не спам.

Задача машинного обучения состоит в том, чтобы из некоторого семейства функций (принимающих на вход параметры, а на выходе выдают результат) найти такую, которая хорошо решает поставленную задачу.

Постановка задачи машинного обучения

Задача: определение спама по входящим звонкам.

Семейство функций(алгоритмы), которые на вход получают аудио звонка и выдают результат: спам или не спам.

Задача машинного обучения состоит в том, чтобы из некоторого семейства функций (принимающих на вход параметры, а на выходе выдают результат) найти такую, которая хорошо решает поставленную задачу.

А также необходимо ввести критерии того, что функция подходит для решения задачи и найти способ как определить эту лучшую функцию.

Основные понятия

Основные понятия

Объект – это предмет изучения, то для чего мы делаем предсказание.

Основные понятия

Объект – это предмет изучения, то для чего мы делаем предсказание.

В нашей задаче - аудиозапись звонка.

Основные понятия

Объект – это предмет изучения, то для чего мы делаем предсказание.

В нашей задаче - аудиозапись звонка.

Признак - характеристика объекта, по которым производится предсказание.

Основные понятия

Объект – это предмет изучения, то для чего мы делаем предсказание.

В нашей задаче - аудиозапись звонка.

Признак - характеристика объекта, по которым производится предсказание.

Целевая переменная – это то что предсказываем.

Основные понятия

Объект – это предмет изучения, то для чего мы делаем предсказание.

В нашей задаче - аудиозапись звонка.

Признак - характеристика объекта, по которым производится предсказание.

Целевая переменная – это то что предсказываем.

В нашей задаче - относится ли звонок в спаму или нет.

Основные понятия

Объект – это предмет изучения, то для чего мы делаем предсказание.

В нашей задаче - аудиозапись звонка.

Признак - характеристика объекта, по которым производится предсказание.

Целевая переменная – это то что предсказываем.

В нашей задаче - относится ли звонок в спаму или нет.

Обучающая выборка - это набор объектов для которых мы знаем правильные ответы и откуда мы будем извлекать закономерности.

Основные понятия

Объект – это предмет изучения, то для чего мы делаем предсказание.

В нашей задаче - аудиозапись звонка.

Признак - характеристика объекта, по которым производится предсказание.

Целевая переменная – это то что предсказываем.

В нашей задаче - относится ли звонок в спаму или нет.

Обучающая выборка - это набор объектов для которых мы знаем правильные ответы и откуда мы будем извлекать закономерности.

В нашей задаче – набор аудиозаписей, промаркированных относятся ли они к спаму или нет.

Основные понятия

Объект – это предмет изучения, то для чего мы делаем предсказание.

В нашей задаче - аудиозапись звонка.

Признак - характеристика объекта, по которым производится предсказание.

Целевая переменная – это то что предсказываем.

В нашей задаче - относится ли звонок в спаму или нет.

Обучающая выборка - это набор объектов для которых мы знаем правильные ответы и откуда мы будем извлекать закономерности.

В нашей задаче – набор аудиозаписей, промаркированных относятся ли они к спаму или нет.

Модель, алгоритм – то что предсказывает. Функция, которая переводит множество объектов в множество ответов.

Основные понятия

x – объект

Основные понятия

x – объект

x_1, x_2, \dots, x_l - все объекты в выборке

Основные понятия

x – объект

x_1, x_2, \dots, x_l - все объекты в выборке

$x_i = (x_{i1}, \dots, x_{id})$ – признаки объекта

d – число признаков

Основные понятия

x – объект

x_1, x_2, \dots, x_l - все объекты в выборке

$x_i = (x_{i1}, \dots, x_{id})$ – признаки объекта

d – число признаков

y – целевая переменная

y_1, y_2, \dots, y_l - все ответы в выборке

Основные понятия

x – объект

x_1, x_2, \dots, x_l - все объекты в выборке

$x_i = (x_{i1}, \dots, x_{id})$ – признаки объекта

d – число признаков

y – целевая переменная

y_1, y_2, \dots, y_l - все ответы в выборке

X – обучающая выборка

Основные понятия

x – объект

x_1, x_2, \dots, x_l - все объекты в выборке

$x_i = (x_{i1}, \dots, x_{id})$ – признаки объекта

d – число признаков

y – целевая переменная

y_1, y_2, \dots, y_l - все ответы в выборке

X – обучающая выборка

$X = \{(x_1, y_1), \dots, (x_l, y_l)\}$

l – длина обучающей выборки

Основные понятия

x – объект

x_1, x_2, \dots, x_l - все объекты в выборке

$x_i = (x_{i1}, \dots, x_{id})$ – признаки объекта

d – число признаков

y – целевая переменная

y_1, y_2, \dots, y_l - все ответы в выборке

X – обучающая выборка

$X = \{(x_1, y_1), \dots, (x_l, y_l)\}$

l – длина обучающей выборки

$X = \begin{pmatrix} x_{11} & \cdots & x_{1d} & y_1 \\ \vdots & \ddots & \vdots & | & \vdots \\ x_{l1} & \cdots & x_{ld} & y_l \end{pmatrix}$ - матрица объекты-признаки

Основные понятия

x – объект

x_1, x_2, \dots, x_l – все объекты в выборке

$x_i = (x_{i1}, \dots, x_{id})$ – признаки объекта

d – число признаков

y – целевая переменная

y_1, y_2, \dots, y_l – все ответы в выборке

X – обучающая выборка

$X = \{(x_1, y_1), \dots, (x_l, y_l)\}$

l – длина обучающей выборки

$X = \begin{pmatrix} x_{11} & \cdots & x_{1d} & y_1 \\ \vdots & \ddots & \vdots & \vdots \\ x_{l1} & \cdots & x_{ld} & y_l \end{pmatrix}$ – матрица объекты-признаки

$a: x_1, x_2, \dots, x_l \rightarrow y_1, y_2, \dots, y_l$ – алгоритм

Типы признаков

Типы признаков

- бинарные

Типы признаков

- бинарные
- числовые

Типы признаков

- бинарные
- числовые
- категориальные

Типы признаков

- бинарные
- числовые
- категориальные
- порядковые

Типы признаков

- бинарные
- числовые
- категориальные
- порядковые
- множественные

Типы признаков

- бинарные
- числовые
- категориальные
- порядковые
- множественные
- сложные (текст, картинка, аудио)

Типы задач машинного обучения

Типы задач машинного обучения

Обучение с учителем

Типы задач машинного обучения

Обучение с учителем

- регрессия

Типы задач машинного обучения

Обучение с учителем

- регрессия
- классификация

Типы задач машинного обучения

Обучение с учителем

- регрессия
- классификация
- бинарная классификация

Типы задач машинного обучения

Обучение с учителем

- регрессия
- классификация
- бинарная классификация
- многоклассовая классификация

Типы задач машинного обучения

Обучение с учителем

- регрессия
- классификация
- бинарная классификация
- многоклассовая классификация
- ранжирование

Типы задач машинного обучения

Обучение с учителем

- регрессия
- классификация
- бинарная классификация
- многоклассовая классификация
- ранжирование

Обучение без учителя

Типы задач машинного обучения

Обучение с учителем

- регрессия
- классификация
- бинарная классификация
- многоклассовая классификация
- ранжирование

Обучение без учителя

- кластеризация

Типы задач машинного обучения

Обучение с учителем

- регрессия
- классификация
- бинарная классификация
- многоклассовая классификация
- ранжирование

Обучение без учителя

- кластеризация
- понижение размерности

Типы задач машинного обучения

Обучение с учителем

- регрессия
- классификация
- бинарная классификация
- многоклассовая классификация
- ранжирование

Обучение без учителя

- кластеризация
- понижение размерности
- определение аномалий

Типы задач машинного обучения

Задача регрессии:

$$y_1, y_2, \dots, y_l \in R$$

Типы задач машинного

обучения

Задача регрессии:

$$y_1, y_2, \dots, y_l \in \mathbb{R}$$

Задача классификации:

$$y_1, y_2, \dots, y_l \in \{1, 2, \dots, k\}$$

Линейные модели

$$a(x) = w_0 + w_1x_1 + \dots + w_dx_d$$

Линейные модели

$$a(x) = w_0 + w_1x_1 + \dots + w_dx_d$$

Функция потерь:

$$L(y, z) \rightarrow R_+$$

z – прогноз модели

Линейные модели

$$a(x) = w_0 + w_1x_1 + \dots + w_dx_d$$

Функция потерь:

$$L(y, z) \rightarrow R_+$$

z – прогноз модели

Функционал ошибки:

$$Q(a, X) = \frac{1}{l} \sum_{i=1}^l L(y_i, a(x_i))$$

Линейные модели

$$a(x) = w_0 + w_1x_1 + \dots + w_dx_d$$

Функция потерь:

$$L(y, z) \rightarrow R_+$$

z – прогноз модели

Функционал ошибки:

$$Q(a, X) = \frac{1}{l} \sum_{i=1}^l L(y_i, a(x_i))$$

$$Q(a, X) \rightarrow \min_a$$

Линейные модели

Задача регрессии

$$L(y, z) = (y - z)^2$$

Линейные модели

Задача регрессии

$$L(y, z) = (y - z)^2$$

Задача классификации

$$L(y, z) = [y \neq z]$$

Линейная регрессия

Задача регрессии

$$a(x) = w_0 + w_1x_1 + \dots + w_dx_d = w_0 + \sum_{j=1}^d w_jx_j$$

Линейная регрессия

Задача регрессии

$$a(x) = w_0 + w_1x_1 + \dots + w_dx_d = w_0 + \sum_{j=1}^d w_jx_j$$

w_0, \dots, w_d - веса

Линейная регрессия

Задача регрессии

$$a(x) = w_0 + w_1x_1 + \dots + w_dx_d = w_0 + \sum_{j=1}^d w_jx_j$$

w_0, \dots, w_d - веса

w_0 - свободный коэффициент (смещение)

Линейная регрессия

Задача регрессии

$$a(x) = w_0 + w_1x_1 + \dots + w_dx_d = w_0 + \sum_{j=1}^d w_jx_j$$

w_0, \dots, w_d - веса

w_0 - свободный коэффициент (смещение)

$$a(x) = w_0 + \langle w, x \rangle$$

Линейная регрессия

Задача регрессии

$$a(x) = w_0 + w_1x_1 + \dots + w_dx_d = w_0 + \sum_{j=1}^d w_jx_j$$

w_0, \dots, w_d - веса

w_0 - свободный коэффициент (смещение)

$$a(x) = w_0 + \langle w, x \rangle$$

$$a(x) = \langle w, x \rangle$$

Подготовка признаков

1) категориальные признаки

Подготовка признаков

1) категориальные признаки

one hot encoding

Подготовка признаков

1) категориальные признаки

one hot encoding

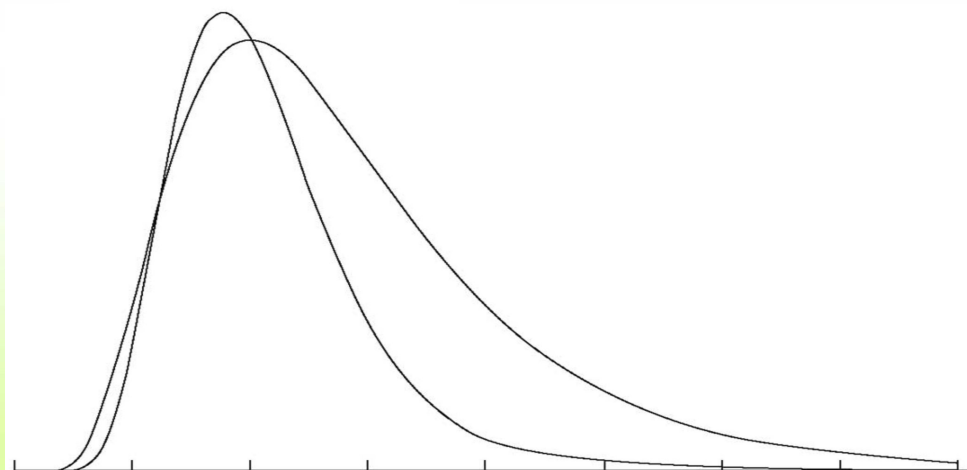
2) нелинейные признаки

Подготовка признаков

1) категориальные признаки
one hot encoding

2) нелинейные признаки

бинаризация



Измерение ошибки в задачах регрессии

Квадратичная функция потерь

$$L(y, z) = (y - z)^2$$

$$MSE(a, X) = \frac{1}{l} \sum_{i=1}^l (a(x_i) - y_i)^2$$

$$RMSE(a, X) = \sqrt{MSE(a, X)}$$

Измерение ошибки в задачах регрессии

Коэффициент детерминации

$$R^2(a, X) = 1 - \frac{\sum_{i=1}^l (a(x_i) - y_i)^2}{\sum_{i=1}^l (y_i - \bar{y})^2}$$

\bar{y} - среднее значение

$R^2 = 1$ – наилучшая модель

$R^2 = 0$

Измерение ошибки в задачах регрессии

Средняя абсолютная ошибка

$$L(y, z) = |y - z|$$

$$MAE(a, X) = \frac{1}{l} \sum_{i=1}^l |a(x_i) - y_i|$$

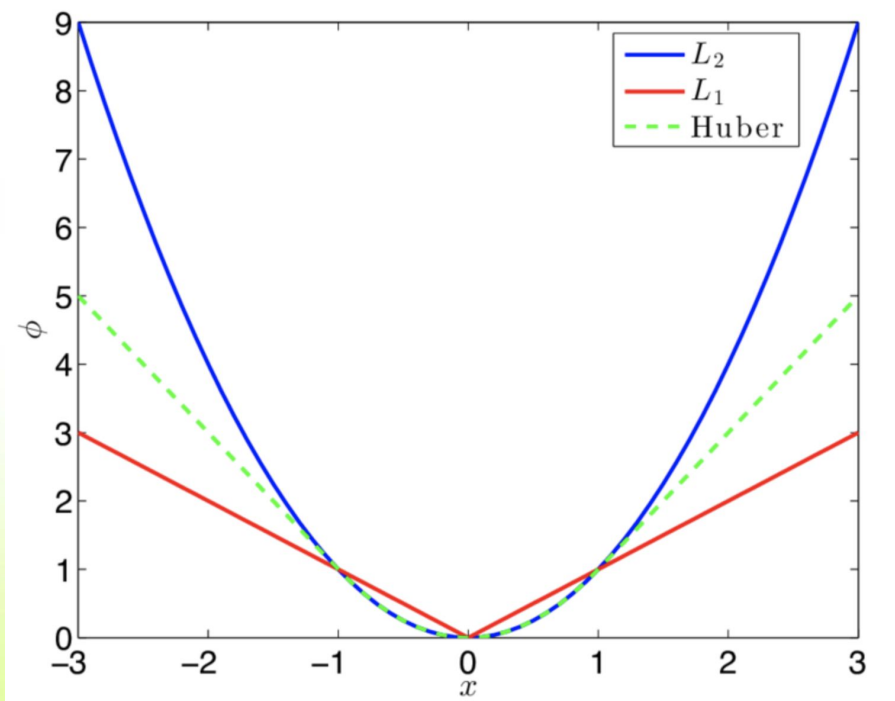
Измерение ошибки в задачах регрессии

y	a(x)	MSE	MAE
1	2	1	1
100	2	9604	98
1	1	0	0
100	3	9409	97

MAE более устойчива к выбросам

Измерение ошибки в задачах регрессии

Функция Хубера (Huber loss)



Измерение ошибки в задачах регрессии

$$L(y, z) = (\log(y + 1) - \log(z + 1))^2$$

$$MSLE(a, X) = \frac{1}{l} \sum_{i=1}^l (\log(y_i + 1) - \log(a(x_i) + 1))^2$$

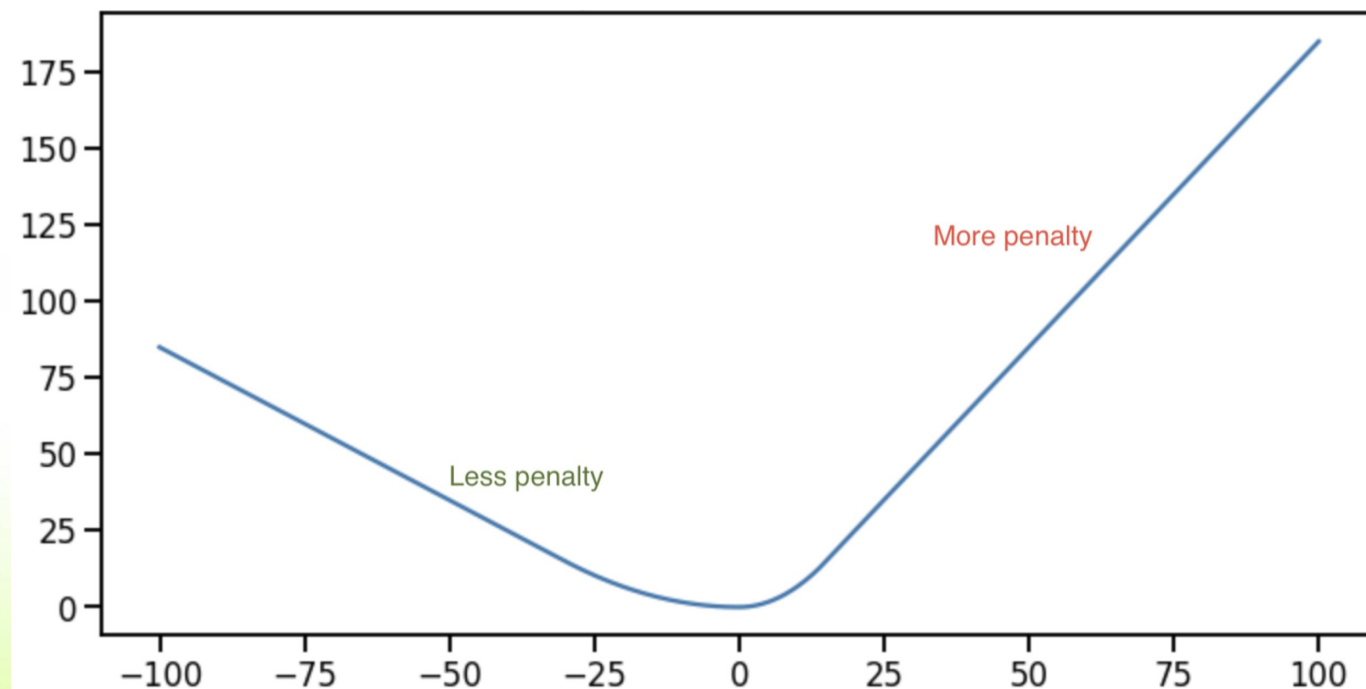
Измерение ошибки в задачах регрессии

$$L(y, z) = \left| \frac{y - z}{y} \right|$$

$$MAPE(a, X) = \frac{1}{l} \sum_{i=1}^l \left| \frac{y_i - a(x_i)}{y_i} \right|$$

Измерение ошибки в задачах регрессии

Несимметричные функции потерь



Обобщающая способность модели

Обобщающая способность модели

Обобщающая способность – способность на новых объектах давать хорошее качество.

Обобщающая способность модели

Обобщающая способность – способность на новых объектах давать хорошее качество.

Недообучение – не удастся найти функцию, которая хорошо описывает данные.

Нужно использовать более сложную функцию.

Обобщающая способность модели

Обобщающая способность – способность на новых объектах давать хорошее качество.

Недообучение – не удастся найти функцию, которая хорошо описывает данные.

Нужно использовать более сложную функцию.

Переобучение – случай когда найденная функция слишком сложная.

Борьба с переобучением



Борьба с переобучением

- валидация модели на отложенной выборке

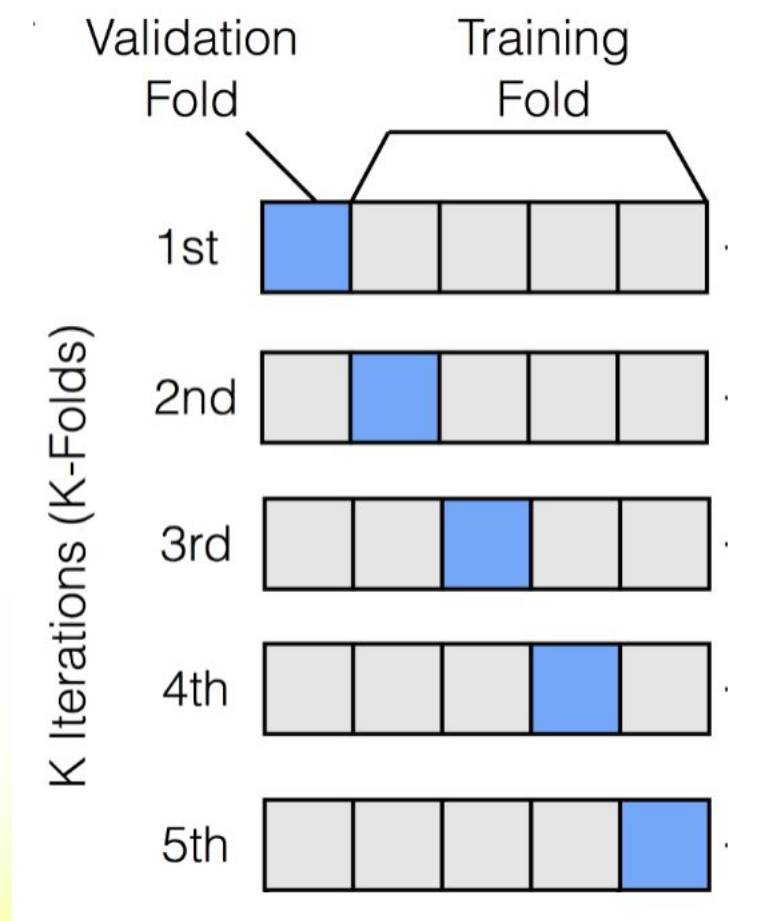
Борьба с переобучением

- валидация модели на отложенной выборке
- кросс-валидация

Борьба с переобучением

- валидация модели на отложенной выборке
- кросс-валидация

leave-one-out



Борьба с переобучением

- валидация модели на отложенной выборке
- кросс-валидация
- регуляризация

Регуляризация

Если посмотрим на веса переобученной модели, то они будут большими

Если большие веса, то при небольших изменениях признаков, результат меняется сильно.

Регуляризация

L2-регуляризация (Ridge)

$Q(w, X)$ - функционал

$$Q(w, X) + \lambda \|w\|_2^2 \rightarrow \min_w$$

$$Q(w, X) + \lambda \sum_{j=1}^d w_j^2 \rightarrow \min_w$$

λ – коэффициент регуляризации

Регуляризация

L1-регуляризация (Lasso)

$$Q(w, X) + \lambda \|w\|_1 \rightarrow \min_w$$

$$Q(w, X) + \lambda \sum_{j=1}^d |w_j| \rightarrow \min_w$$

Регуляризация

L1 регуляризация обнуляет веса, это можно использовать для прореживания модели.

Для чего это нужно:

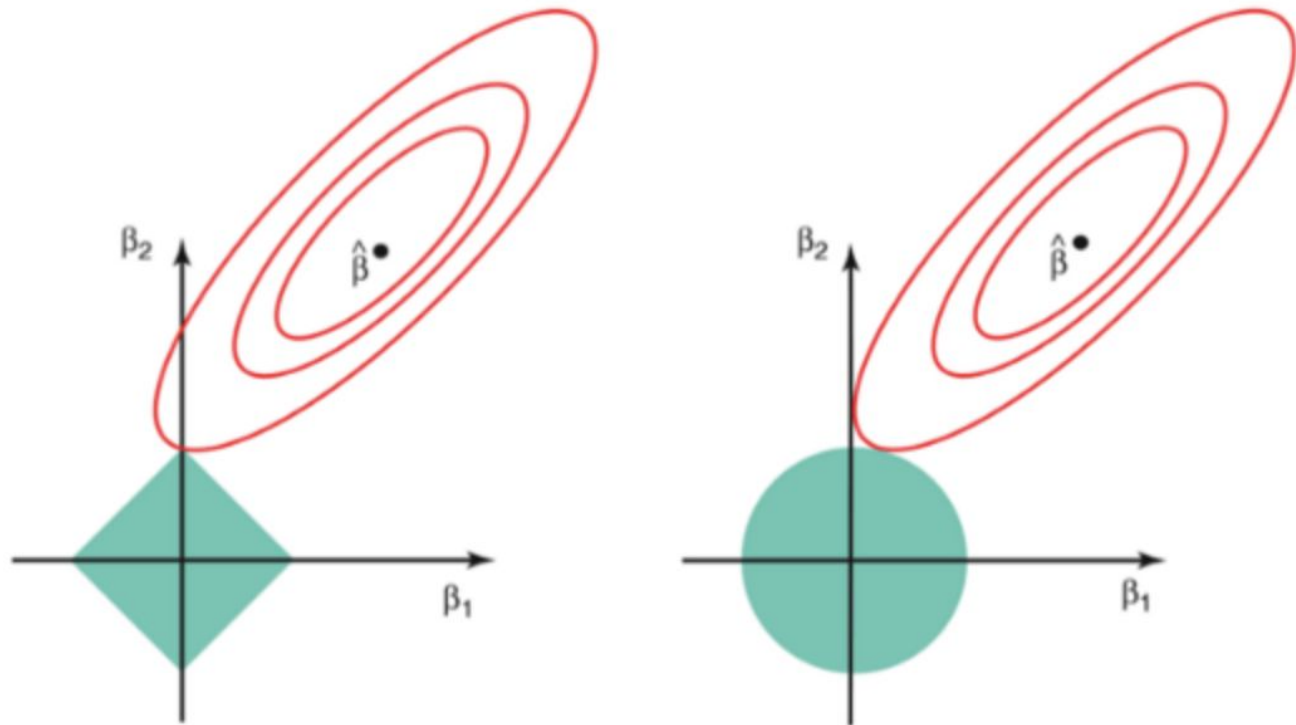
- ускорение модели
- мы знаем что есть лишние признаки(линейно зависимые, шумные)
- данных меньше чем признаков

Регуляризация

L1-регуляризация (Lasso)

$$Q(w, X) + \lambda \|w\|_1 \rightarrow \min_w$$

$$\begin{cases} Q(w, X) \rightarrow \min_w \\ \|w\|_1 \leq C \end{cases}$$



Credit : An Introduction to Statistical Learning by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani

Обучение линейной модели

Обучение линейной модели

$$\frac{1}{l} \sum_{i=1}^l (\langle w, x_i \rangle - y_i)^2 \rightarrow \min_w$$

Обучение линейной модели

$$\frac{1}{l} \sum_{i=1}^l (\langle w, x_i \rangle - y_i)^2 \rightarrow \min_w$$

$$MSE = \frac{1}{l} \|Xw - y\|_2^2 \rightarrow \min_w$$

Обучение линейной модели

$$\frac{1}{l} \sum_{i=1}^l (\langle w, x_i \rangle - y_i)^2 \rightarrow \min_w$$

$$MSE = \frac{1}{l} \|Xw - y\|_2^2 \rightarrow \min_w$$

$$\nabla_w MSE = 0$$

Обучение линейной модели

$$\frac{1}{l} \sum_{i=1}^l (< w, x_i > - y_i)^2 \rightarrow \min_w$$

$$MSE = \frac{1}{l} \|Xw - y\|_2^2 \rightarrow \min_w$$

$$\nabla_w MSE = 0$$

$$w = (X^T X)^{-1} X^T y$$

Обучение линейной модели

$$\frac{1}{l} \sum_{i=1}^l (< w, x_i > - y_i)^2 \rightarrow \min_w$$

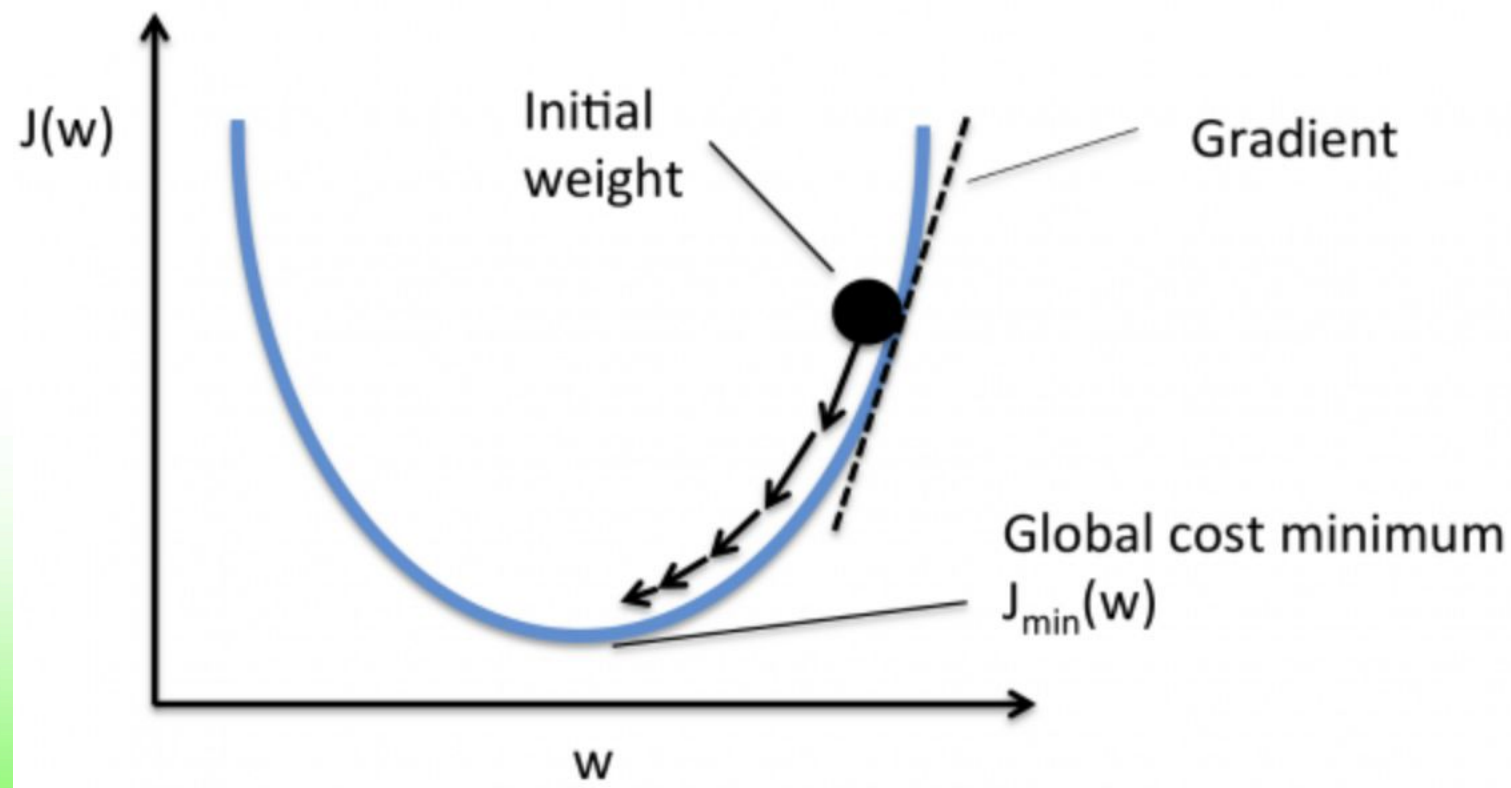
$$MSE = \frac{1}{l} \|Xw - y\|_2^2 \rightarrow \min_w$$

$$\nabla_w MSE = 0$$

$$w = (X^T X)^{-1} X^T y$$

X - матрица полного ранга (нет линейно зависимых признаков)

Градиентный спуск



Градиентный спуск

$w^{(0)}$ - начальное приближение весов

$\nabla_w Q(w)$ - градиент

$-\nabla_w Q(w)$ - антиградиент

$$w^{(k)} = w^{(k-1)} - \mu_k \nabla_w Q(w^{(k-1)})$$

μ – длина шага

Градиентный спуск

Критерии остановки

- когда ошибка на тесте перестает уменьшаться
- когда ошибка на обучающей выборке не поменялась
- когда вектор весов не сильно меняется

Оптимизации градиентного спуска

$$Q(w, X) = \frac{1}{l} \sum_{i=1}^l (< w, x_i > - y_i)^2$$

$$Q(w, X) = \frac{1}{l} \sum_{i=1}^l q_i(w)$$

$$\nabla_w Q(w, X) = \frac{1}{l} \sum_{i=1}^l \nabla_w q_i(w) \text{ - полный градиент}$$

Оптимизации градиентного спуска

Стохастический градиентный спуск (Stochastic gradient descent, SGD)

$$\nabla_w Q(w) \approx \nabla_w q_i(w)$$

$$w^{(k)} = w^{(k-1)} - \mu_k \nabla_w q_{i_k}(w^{(k-1)})$$

Оптимизации градиентного спуска

Mini-batch gradient descent

$$\nabla_w Q(w) \approx \frac{1}{n} \sum_{j=1}^n \nabla_w q_i(w)$$

n - размер батча

$$w^{(k)} = w^{(k-1)} - \mu_k \frac{1}{n} \sum_{j=1}^n \nabla_w q_{i_k}(w^{(k-1)})$$

Вопросы