Введение в искусственный интеллект. Машинное обучение

Тема: Линейная регрессия

Бабин Д.Н., Иванов И.Е.

кафедра Математической Теории Интеллектуальных Систем

План лекции

1. Оптимальный байесовский регрессор
2. Линейная регрессии с точки зрения ML и MAP оценивания
3. Регуляризация: гребневая регрессия, LASSO, ElasticNet
4. Метрики качества для задачи регрессии

Напоминание: вероятностная постановка задач машинного обучения

Предположения

Пусть известно совместное распределение $p(x, y)$ на $X \times Y$
Пусть задана функция потерь $L(a(x), y)$

Определение

Средняя величина потерь для алгоритма $a(x)$

$$R(a) = \iint L(a(x), y) dP(x, y) = \iint L(a(x), y) p(x, y) dx dy$$

Задача

Найти такой $a^*(x)$, что $a^*(x) = \arg\min_a R(a)$.

Будем называть модель a^* оптимальной и R^* — значение оптимального среднего риска.

Квадратичная функция потерь

Теорема

Если $L(a(x), y) = (a(x) - y)^2$, то величина средних потерь минимальна при

Квадратичная функция потерь

Теорема

Если $L(a(x), y) = (a(x) - y)^2$, то величина средних потерь минимальна при

$$a^* = E(y|x)$$

Квадратичная функция потерь

Теорема

Если $L(a(x), y) = (a(x) - y)^2$, то величина средних потерь минимальна при

$$a^* = E(y|x)$$

Лемма

$$E\left((y - a(x))^2|x\right) = E\left((y - E(y|x))^2|x\right) + E\left((a(x) - E(y|x))^2|x\right)$$

Квадратичная функция потерь

Теорема

Если $L(a(x), y) = (a(x) - y)^2$, то величина средних потерь минимальна при

$$a^* = E(y|x)$$

Лемма

$$E\left((y - a(x))^2|x\right) = E\left((y - E(y|x))^2|x\right) + E\left((a(x) - E(y|x))^2|x\right)$$

Доказательство

$$E\left((y - a(x))^2|x\right) = E\left((y - E(y|x) + E(y|x) - a(x))^2|x\right) =$$

Квадратичная функция потерь

Теорема

Если $L(a(x), y) = (a(x) - y)^2$, то величина средних потерь минимальна при

$$a^* = E(y|x)$$

Лемма

$$E\left((y - a(x))^2|x\right) = E\left((y - E(y|x))^2|x\right) + E\left((a(x) - E(y|x))^2|x\right)$$

Доказательство

$$E\left((y - a(x))^2|x\right) = E\left((y - E(y|x) + E(y|x) - a(x))^2|x\right) =$$
$$= E\left((y - E(y|x))^2|x\right) + E\left((a(x) - E(y|x))^2|x\right) - 2E\left(y - E(y|x)|x\right)E\left(a(x) - E(y|x)|x\right)$$

Квадратичная функция потерь

Теорема

Если $L(a(x), y) = (a(x) - y)^2$, то величина средних потерь минимальна при

$$a^* = E(y|x)$$

Лемма

$$E\left((y - a(x))^2|x\right) = E\left((y - E(y|x))^2|x\right) + E\left((a(x) - E(y|x))^2|x\right)$$

Доказательство

$$E\left((y - a(x))^2|x\right) = E\left((y - E(y|x) + E(y|x) - a(x))^2|x\right) =$$
$$= E\left((y - E(y|x))^2|x\right) + E\left((a(x) - E(y|x))^2|x\right) - 2E\left(y - E(y|x)|x\right)E\left(a(x) - E(y|x)|x\right)$$

Последнее слагаемое равно нулю, так как

$$E\left(y - E(y|x)|x\right) = E(y|x) - E(E(y|x)|x) = E(y|x) - E(y|x) = 0.$$

Квадратичная функция потерь

Теорема

Если $L(a(x), y) = (a(x) - y)^2$, то величина средних потерь минимальна при

$$a^* = E(y|x)$$

Квадратичная функция потерь

Теорема

Если $L(a(x), y) = (a(x) - y)^2$, то величина средних потерь минимальна при

$$a^* = E(y|x)$$

Доказательство

$R(a) = \iint L(a(x), y)p(x, y)dydx = \iint (a(x) - y)^2 p(x, y)dydx =$

Квадратичная функция потерь

Теорема

Если $L(a(x), y) = (a(x) - y)^2$, то величина средних потерь минимальна при

$$a^* = E(y|x)$$

Доказательство

$R(a) = \iint L(a(x), y)p(x, y)dydx = \iint (a(x) - y)^2 p(x, y)dydx =$
$= \iint (a(x) - y)^2 p(y|x)dyp(x)dx = \int E((y - a(x))^2|x)p(x)dx$

Квадратичная функция потерь

Теорема

Если $L(a(x), y) = (a(x) - y)^2$, то величина средних потерь минимальна при

$$a^* = E(y|x)$$

Доказательство

$R(a) = \int\int L(a(x), y)p(x, y)dydx = \int\int(a(x) - y)^2 p(x, y)dydx =$
$= \int\int(a(x) - y)^2 p(y|x)dyp(x)dx = \int E((y - a(x))^2|x)p(x)dx$ Применяя лемму, получаем:
$R(a) = \int E((y - a(x))^2|x)p(x)dx = \int E\left((y - E(y|x))^2|x\right) p(x)dx +$
$\int E\left((a(x) - E(y|x))^2|x\right) p(x)dx \geq \int E\left((y - E(y|x))^2|x\right) p(x)dx$, что и требовалось доказать.

Время для вопросов

Постановка задачи: линейная регрессия

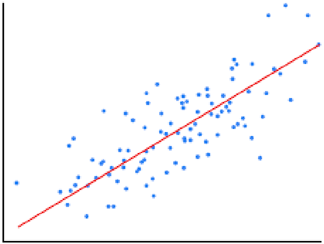
Дано

$$p(y_i|x_i) \sim w^T x_i + \varepsilon_i \sim N(w^T x_i, \sigma^2),$$

для $i = 1..,\ell$, где $w \in R^{n+1}$, $\varepsilon_i \sim N(0, \sigma^2)$

Задача

Найти w



Напоминание: два вида оценивания параметров

Принцип максимального правдоподобия

$$w_{ML} = \arg\max_{w} \; p(y|w, x)$$

Напоминание: два вида оценивания параметров

Принцип максимального правдоподобия

$$w_{ML} = \arg\max_{w} \; p(y|w, x)$$

Принцип максимума апостериорной вероятности

$$w_{MAP} = \arg\max_{w} \; p(w|x, y)$$

Оценка максимального правдободобия

$$w_{ML} = \arg\max_{w} \; p(y|w, x)$$

Оценка максимального правдободобия

$$w_{ML} = \arg\max_w \; p(y|w, x)$$

$$w_{ML} = \arg\max_w \; \prod_i p(y_i|w, x_i)$$

Оценка максимального правдободобия

$$w_{ML} = \arg\max_{w} \; p(y|w, x)$$

$$w_{ML} = \arg\max_{w} \prod_{i} p(y_i|w, x_i)$$

$$p(y_i|w, x_i) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(y_i - w^T x_i)^2}{2\sigma^2}\right)$$

Оценка максимального правдобододобия

$$w_{ML} = \arg\max_{w} \; p(y|w,x)$$

$$w_{ML} = \arg\max_{w} \prod_{i} p(y_i|w,x_i)$$

$$p(y_i|w,x_i) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(y_i - w^T x_i)^2}{2\sigma^2}\right)$$

$$w_{ML} = \arg\max_{w} \prod_{i} \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(y_i - w^T x_i)^2}{2\sigma^2}\right) = \arg\max_{w} \sum_{i} -\frac{(y_i - w^T x_i)^2}{2\sigma^2}$$

Оценка максимального правдободобия

$$w_{ML} = \arg\max_{w} \; p(y|w, x)$$

$$w_{ML} = \arg\max_{w} \; \prod_i p(y_i|w, x_i)$$

$$p(y_i|w, x_i) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(y_i - w^T x_i)^2}{2\sigma^2}\right)$$

$$w_{ML} = \arg\max_{w} \; \prod_i \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(y_i - w^T x_i)^2}{2\sigma^2}\right) = \arg\max_{w} \sum_i -\frac{(y_i - w^T x_i)^2}{2\sigma^2}$$

$$w_{ML} = \arg\min_{w} \; \sum_i (y_i - w^T x_i)^2$$

Метод наименьших квадратов

Постановка задачи и допущения

- $X = \mathbb{R}^n$, $Y = \mathbb{R}$

Метод наименьших квадратов

Постановка задачи и допущения

- $X = \mathbb{R}^n$, $Y = \mathbb{R}$
- $a(x) = f_w(x) = w_0 + w_1x_1 + w_2x_2 + \ldots + w_nx_n$, где $w = (w_0, w_1, \ldots, w_n)^T \in \mathbb{R}^{n+1}$ — параметры модели.

Метод наименьших квадратов

Постановка задачи и допущения

- $X = \mathbb{R}^n$, $Y = \mathbb{R}$
- $a(x) = f_w(x) = w_0 + w_1x_1 + w_2x_2 + \ldots + w_nx_n$, где $w = (w_0, w_1, \ldots, w_n)^T \in \mathbb{R}^{n+1}$ — параметры модели.
- Удобно писать в векторном виде

$$a(x) = w^T \cdot x,$$

где $x = (1, x^1, \ldots, x^n)^T \in \mathbb{R}^{n+1}$.

Метод наименьших квадратов

Постановка задачи и допущения

- $X = \mathbb{R}^n$, $Y = \mathbb{R}$
- $a(x) = f_w(x) = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_n x_n$, где $w = (w_0, w_1, \ldots, w_n)^T \in \mathbb{R}^{n+1}$ — параметры модели.
- Удобно писать в векторном виде

$$a(x) = w^T \cdot x,$$

где $x = (1, x^1, \ldots, x^n)^T \in \mathbb{R}^{n+1}$.

Метод наименьших квадратов

- $L(w, X_{train}) = MSE(w, X_{train}) = \frac{1}{\ell} \sum\limits_i (w^T \cdot x^{(i)} - y_i)^2$ — функция потерь

Метод наименьших квадратов

Постановка задачи и допущения

- $X = \mathbb{R}^n$, $Y = \mathbb{R}$
- $a(x) = f_w(x) = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_n x_n$, где $w = (w_0, w_1, \ldots, w_n)^T \in \mathbb{R}^{n+1}$ — параметры модели.
- Удобно писать в векторном виде

$$a(x) = w^T \cdot x,$$

где $x = (1, x^1, \ldots, x^n)^T \in \mathbb{R}^{n+1}$.

Метод наименьших квадратов

- $L(w, X_{train}) = MSE(w, X_{train}) = \frac{1}{\ell} \sum_i (w^T \cdot x^{(i)} - y_i)^2$ — функция потерь
- Задача найти $\hat{w} = \arg\min_w (L(w, X_{train}))$

Аналитическое решение

Теорема

Решением задачи $\arg\min\limits_{w}(\sum\limits_{i=1}^{\ell}(w^T \cdot x_i - y_i)^2)$ является $\hat{w} = (X^T X)^{-1} \cdot X^T \cdot y$, где $X_{i,j} = x_i^j$, $y = (y_1, ..., y_\ell)$.

Аналитическое решение

Теорема

Решением задачи $\arg\min\limits_w (\sum\limits_{i=1}^{\ell}(w^T \cdot x_i - y_i)^2)$ является $\hat{w} = (X^T X)^{-1} \cdot X^T \cdot y$, где $X_{i,j} = x_i^j$, $y = (y_1, ..., y_\ell)$.

Доказательство

Запишем задачу в векторном виде $||Xw - y||^2 \to \min\limits_w$. Необходимое условие минимума в матричном виде имеет вид:

$$\frac{\partial}{\partial w}||Xw-y||^2 = \frac{\partial}{\partial w}\Big((Xw - y)^T \cdot (Xw - y)\Big) = \frac{\partial}{\partial w}\Big((Xw)^T Xw - (Xw)^T y - y^T Xw + y^T y\Big) =$$

$$\frac{\partial}{\partial w}\Big(w^T X^T Xw - w^T X^T y - y^T Xw + y^T y\Big) = \frac{\partial}{\partial w} w^T (X^T X) w - 2\frac{\partial}{\partial w}(X^T y)^T w = 0$$

Леммы

Определение

Пусть $w = (w_1, ..., w_n)$ — вектор столбец, а $z = z(w_1, ..., w_n)$. Тогда определим

$$\frac{\partial z}{\partial w} := (\frac{\partial z}{\partial w_1}, ..., \frac{\partial z}{\partial w_n})^T$$

Лемма 1

$$\frac{\partial}{\partial x} x^T a = a$$

Лемма 2

$$\frac{\partial}{\partial x} x^T A x = (A + A^T) x$$

Аналитическое решение

Теорема

Решением задачи $\arg\min_w (\sum_{i=1}^{\ell} (w^T \cdot x_i - y_i)^2)$ является $\hat{w} = (X^T X)^{-1} \cdot X^T \cdot y$, где $X_{i,j} = x_i^j$, $y = (y_1, ..., y_\ell)$.

Продолжение доказательтва

Необходимое условие минимума в матричном виде имеет вид:

$$\frac{\partial}{\partial w} ||Xw - y||^2 = \frac{\partial}{\partial w} w^T (X^T X) w - 2 \frac{\partial}{\partial w} (X^T y)^T w =$$

Далее применяем леммы и приравниваем к нулю:

$$= 2X^T X w - 2X^T y = 0,$$

откуда получаем $w = (X^T X)^{-1} \cdot X^T \cdot y$, что и требовалось доказать.

Полиномиальная регрессия

Идея

Можно генерировать новые признаки на основе уже имеющихся, применяя нелинейные функции

Полиномиальная регрессия

Идея

Можно генерировать новые признаки на основе уже имеющихся, применяя нелинейные функции

Примеры преобразований

- Возведение в степень
- Попарные произведения
- Квадратный корень
- Логарифм
- Экспонента

Преимущества и недостатки линейной регрессии

Преимущества и недостатки линейной регрессии

Преимущества

- Простой алгоритм, вычислительно не сложный
- Линейная регрессия хорошо интерпретируемая модель
- Несмотря на свою простоту может описывать довольно сложные зависимости (например, полиномиальные)

Преимущества и недостатки линейной регрессии

Преимущества

- Простой алгоритм, вычислительно не сложный
- Линейная регрессия хорошо интерпретируемая модель
- Несмотря на свою простоту может описывать довольно сложные зависимости (например, полиномиальные)

Недостатки

- Алгоритм предполагает, что все признаки числовые
- Алгоритм предполагает, что данные распределены нормально, что не всегда так
- Алгоритм сильно чувствителен к выбросам

Время для вопросов

Метод максимизации апостериорной вероятности

$$w_{MAP} = \arg\max_{w} \; p(w | x_1, ...x_\ell, y_1, ..., y_\ell)$$

Метод максимизации апостериорной вероятности

$$w_{MAP} = \arg\max_{w} \; p(w|x_1, ...x_\ell, y_1, ..., y_\ell)$$

$$w_{MAP} = \arg\max_{w} \; \prod_{i} p(y_i|x_i, w)p(w)$$

Метод максимизации апостериорной вероятности

$$w_{MAP} = \arg\max_{w} \; p(w|x_1, ...x_\ell, y_1, ..., y_\ell)$$

$$w_{MAP} = \arg\max_{w} \; \prod_i p(y_i|x_i, w)p(w)$$

$$w_{MAP} = \arg\max_{w} \; \sum_i \ln p(y_i|x_i, w) + \ln p(w)$$

Метод максимизации апостериорной вероятности

$$w_{MAP} = \arg\max_{w} \; p(w|x_1, ...x_\ell, y_1, ..., y_\ell)$$

$$w_{MAP} = \arg\max_{w} \; \prod_{i} p(y_i|x_i, w)p(w)$$

$$w_{MAP} = \arg\max_{w} \; \sum_{i} \ln p(y_i|x_i, w) + \ln p(w)$$

$$w_{MAP} = \arg\max_{w} \; \sum_{i} -\frac{(y_i - w^T x_i)^2}{2\sigma^2} + \ell \ln p(w)$$

Метод максимизации апостериорной вероятности

$$w_{MAP} = \arg\max_{w} \; p(w|x_1, \ldots x_\ell, y_1, \ldots, y_\ell)$$

$$w_{MAP} = \arg\max_{w} \; \prod_i p(y_i|x_i, w)p(w)$$

$$w_{MAP} = \arg\max_{w} \; \sum_i \ln p(y_i|x_i, w) + \ln p(w)$$

$$w_{MAP} = \arg\max_{w} \; \sum_i -\frac{(y_i - w^T x_i)^2}{2\sigma^2} + \ell \ln p(w)$$

$$w_{MAP} = \arg\min_{w} \; \sum_i \frac{(y_i - w^T x_i)^2}{2\sigma^2} - \ell \ln p(w)$$

Метод максимизации апостериорной вероятности

$$w_{MAP} = \arg\max_{w} \; p(w|x_1, \ldots x_\ell, y_1, \ldots, y_\ell)$$

$$w_{MAP} = \arg\max_{w} \; \prod_i p(y_i|x_i, w) p(w)$$

$$w_{MAP} = \arg\max_{w} \; \sum_i \ln p(y_i|x_i, w) + \ln p(w)$$

$$w_{MAP} = \arg\max_{w} \; \sum_i -\frac{(y_i - w^T x_i)^2}{2\sigma^2} + \ell \ln p(w)$$

$$w_{MAP} = \arg\min_{w} \; \sum_i \frac{(y_i - w^T x_i)^2}{2\sigma^2} - \ell \ln p(w)$$

В задаче минимизации появилось дополнительное слагаемое, которое зависит только от априорного распределения на веса w

Метод максимизации апостериорной вероятности

$$w_{MAP} = \arg\min_{w} \sum_{i} \frac{(y_i - w^T x_i)^2}{2\sigma^2} - \ell \ln p(w)$$

Метод максимизации апостериорной вероятности

$$w_{MAP} = \arg\min_{w} \sum_{i} \frac{(y_i - w^T x_i)^2}{2\sigma^2} - \ell \ln p(w)$$

Предположим, что $p(w) \sim N(0, \tau^2)$

$$w_{MAP} = \arg\min_w \sum_i \frac{(y_i - w^T x_i)^2}{2\sigma^2} - \ell \ln p(w)$$

Предположим, что $p(w) \sim N(0, \tau^2)$

$$w_{MAP} = \arg\min_w \sum_i \frac{(y_i - w^T x_i)^2}{2\sigma^2} - \frac{\ell w^T w}{2\tau^2}$$

Метод максимизации апостериорной вероятности

$$w_{MAP} = \arg\min_w \sum_i \frac{(y_i - w^T x_i)^2}{2\sigma^2} - \ell \ln p(w)$$

Предположим, что $p(w) \sim N(0, \tau^2)$

$$w_{MAP} = \arg\min_w \sum_i \frac{(y_i - w^T x_i)^2}{2\sigma^2} - \frac{\ell w^T w}{2\tau^2}$$

$$w_{MAP} = \arg\min_w \frac{1}{\ell} \sum_i \frac{(y_i - w^T x_i)^2}{2\sigma^2} - \frac{1}{2\tau^2} ||w||^2$$

Гребневая регрессия

Гребневая регрессия

L2-регуляризация

- $L(w, X_{train}) = MSE(w, X_{train}) + \frac{\alpha}{2}\sum\limits_{i=0}^{n} w_i^2 = \frac{1}{\ell}\sum\limits_{i}(w^T \cdot x^{(i)} - y_i)^2 + \frac{\alpha}{2}\sum\limits_{i=0}^{n} w_i^2$ — функция потерь

Гребневая регрессия

L2-регуляризация

- $L(w, X_{train}) = MSE(w, X_{train}) + \frac{\alpha}{2}\sum\limits_{i=0}^{n} w_i^2 = \frac{1}{\ell}\sum\limits_{i}(w^T \cdot x^{(i)} - y_i)^2 + \frac{\alpha}{2}\sum\limits_{i=0}^{n} w_i^2$ — функция потерь
- Задача найти $\hat{w} = \arg\min\limits_{w}(L(w, X_{train}))$

Гребневая регрессия

L2-регуляризация

- $L(w, X_{train}) = MSE(w, X_{train}) + \frac{\alpha}{2}\sum_{i=0}^{n} w_i^2 = \frac{1}{\ell}\sum_{i}(w^T \cdot x^{(i)} - y_i)^2 + \frac{\alpha}{2}\sum_{i=0}^{n} w_i^2$ — функция потерь
- Задача найти $\hat{w} = \arg\min_{w}(L(w, X_{train}))$

Теорема

Решением задачи $\arg\min_{w}(\sum_{i=1}^{\ell}(w^T \cdot x^{(i)} - y_i)^2 + \alpha\sum_{i=0}^{n} w_i^2)$ является $\hat{w} = (X^T X + \alpha I_{n+1})^{-1} \cdot X^T \cdot y$, где $X_{i,j} = x_i^j$, $y = (y_1, \ldots, y_\ell)$, I_{n+1} — единичная матрица.

Доказательство теоремы

Доказательство теоремы

Лемма 3

$$\frac{\partial}{\partial x} x^T x = 2x$$

Доказательство теоремы

Лемма 3

$$\frac{\partial}{\partial x} x^T x = 2x$$

Теорема

Решением задачи $\arg\min\limits_{w}(\sum\limits_{i=1}^{\ell}(w^T \cdot x^{(i)} - y_i)^2 + \alpha \sum\limits_{i=0}^{n} w_i^2)$ является $\hat{w} = (X^T X + \alpha I_{n+1})^{-1} \cdot X^T \cdot y$, где $X_{i,j} = x_i^j$, $y = (y_1, ..., y_\ell)$, I_{n+1} — единичная матрица.

Доказательство

Запишем задачу в векторном виде $||Xw - y||^2 + \alpha||w||^2 \to \min\limits_{w}$. Необходимое условие минимума в матричном виде имеет вид:

$$\frac{\partial}{\partial w}\left((Xw - y)^T \cdot (Xw - y) + \alpha w^T w\right) = 2X^T Xw - 2X^T y + 2\alpha w = 0$$

Свойства гребневой регрессии

- Регуляризация не даёт параметрам модели быть слишком большими
- Как правило регуляризация обеспечивает большую обобщающую способность
- Более устойчива к выбросам
- Появился параметр, который можно настравить при помощи кросс-валидации

Свойства гребневой регрессии

- Регуляризация не даёт параметрам модели быть слишком большими
- Как правило регуляризация обеспечивает большую обобщающую способность
- Более устойчива к выбросам
- Появился параметр, который можно настравить при помощи кросс-валидации

Вероятностный смысл параметра α

$\alpha = \frac{1}{\tau^2}$, где τ — среднеквадратическое отклонение априорного распределения на w

LASSO

L1-регуляризация

- $L(w, X_{train}) = MSE(w, X_{train}) + \alpha \sum_{i=0}^{n} |w_i| = \sum_i (w^T \cdot x^{(i)} - y_i)^2 + \alpha \sum_{i=0}^{n} |w_i|$ — функция потерь

L1-регуляризация

- $L(w, X_{train}) = MSE(w, X_{train}) + \alpha \sum_{i=0}^{n} |w_i| = \sum_{i} (w^T \cdot x^{(i)} - y_i)^2 + \alpha \sum_{i=0}^{n} |w_i|$ — функция потерь
- Задача найти $\hat{w} = \arg\min_{w}(L(w, X_{train}))$

LASSO

L1-регуляризация

- $L(w, X_{train}) = MSE(w, X_{train}) + \alpha \sum\limits_{i=0}^{n} |w_i| = \sum\limits_{i} (w^T \cdot x^{(i)} - y_i)^2 + \alpha \sum\limits_{i=0}^{n} |w_i|$ — функция потерь
- Задача найти $\hat{w} = \arg\min\limits_{w} (L(w, X_{train}))$

Свойства

- Эта регуляризация обеспечивает отбор признаков
- Нет аналитического решения

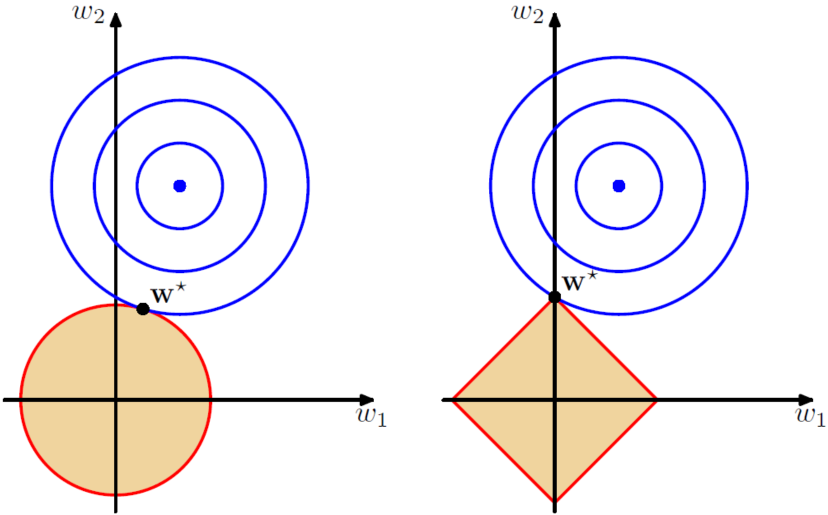
Вероятностная интерпретация LASSO

Вероятностный смысл параметра α

Параметр α — обратно пропорционален среднеквадратичному отклонению априорного распределения на w. В данном случае это распределение Лапласа

$$p(w) = \frac{1}{\tau} exp\left(-\frac{||w||}{2\tau}\right)$$

Интуиция отбора признаков при L1-регуляризации



Elastic Net

L1-регуляризация и L2-регуляризация

- $L(w, X_{train}) = MSE(w, X_{train}) + r\alpha \sum_{i=0}^{n} |w_i| + (1-r)\frac{\alpha}{2} \sum_{i=0}^{n} w_i^2 = \sum_i (w^T \cdot x^{(i)} - y_i)^2 + r\alpha \sum_{i=0}^{n} |w_i| + (1-r)\frac{\alpha}{2} \sum_{i=0}^{n} w_i^2$ — функция потерь

Elastic Net

L1-регуляризация и L2-регуляризация

- $L(w, X_{train}) = MSE(w, X_{train}) + r\alpha \sum\limits_{i=0}^{n} |w_i| + (1-r)\frac{\alpha}{2} \sum\limits_{i=0}^{n} w_i^2 =$
  $\sum\limits_{i} (w^T \cdot x^{(i)} - y_i)^2 + r\alpha \sum\limits_{i=0}^{n} |w_i| + (1-r)\frac{\alpha}{2} \sum\limits_{i=0}^{n} w_i^2$ — функция потерь
- Задача найти $\hat{w} = \arg\min\limits_{w}(L(w, X_{train}))$

Свойства

- Нет аналитического решения
- Совмещает положительные свойства гребневой регрессии и LASSO.

Время для вопросов

Мотивация

- Постановка задачи машинного обучения обычно начинается с определения метрики и фиксирования тестового датасета, на котором эта метрика будет считаться

Метрики качества для задачи регрессии

Мотивация

- Постановка задачи машинного обучения обычно начинается с определения метрики и фиксирования тестового датасета, на котором эта метрика будет считаться
- Неправильно выбранная метрика может затруднить использование модели машинного обучения в жизни и свести на нет усилия команды, разрабатываюшей алгоритм машинного обучения.

Метрики качества для задачи регрессии

Мотивация

- Постановка задачи машинного обучения обычно начинается с определения метрики и фиксирования тестового датасета, на котором эта метрика будет считаться
- Неправильно выбранная метрика может затруднить использование модели машинного обучения в жизни и свести на нет усилия команды, разрабатываюшей алгоритм машинного обучения.
- Как правило заказчик не мыслит в терминах метрик и может объяснить проблему, которую он хочет решить, только бизнес языком

Метрики качества для задачи регрессии

Мотивация

- Постановка задачи машинного обучения обычно начинается с определения метрики и фиксирования тестового датасета, на котором эта метрика будет считаться
- Неправильно выбранная метрика может затруднить использование модели машинного обучения в жизни и свести на нет усилия команды, разрабатываюшей алгоритм машинного обучения.
- Как правило заказчик не мыслит в терминах метрик и может объяснить проблему, которую он хочет решить, только бизнес языком
- Понимание влияния выбора той или иной метрики на бизнес — это ключ к успешной постановки задачи

Метрики качества для задачи регрессии

Mean Square Error

$$MSE = \frac{1}{\ell}\sum_i (y_i - a(x_i))^2$$

Метрики качества для задачи регрессии

Mean Square Error

$$MSE = \frac{1}{\ell}\sum_i (y_i - a(x_i))^2$$

Root Mean Square Error

$$RMSE = \sqrt{\frac{1}{\ell}\sum_i (y_i - a(x_i))^2}$$

Метрики качества для задачи регрессии

Mean Square Error

$$MSE = \frac{1}{\ell} \sum_i (y_i - a(x_i))^2$$

Root Mean Square Error

$$RMSE = \sqrt{\frac{1}{\ell} \sum_i (y_i - a(x_i))^2}$$

Mean Absolute Error

$$MAE = \frac{1}{\ell} \sum_i |y_i - a(x_i)|$$

Метрики качества для задачи регрессии

Max Error

$$ME = max(|y_i - a(x_i)|)$$

Метрики качества для задачи регрессии

Max Error

$$ME = max(|y_i - a(x_i)|)$$

Mean Squared Logarithmic Error

$$MSLE = \frac{1}{\ell} \sum_i (\ln y_i - \ln a(x_i))^2$$

Метрики качества для задачи регрессии

Max Error

$$ME = max(|y_i - a(x_i)|)$$

Mean Squared Logarithmic Error

$$MSLE = \frac{1}{\ell}\sum_i (\ln y_i - \ln a(x_i))^2$$

R^2 score

$$R^2 = 1 - \frac{\sum_i (y_i - a(x_i))^2}{\sum_i (y_i - \bar{y})^2},$$

где $\bar{y} = \frac{1}{\ell}\sum_i y_i$

Время для вопросов

Заключение

- Линейная регрессия — простая, хорошо интерпретируемая модель, не устойчивая к выбросам
- Имеет наглядную вероятностную интерпретацию
- Регуляризация — отличный способ борьбы с переобучением и шумом в данных

Время для вопросов