Основы программирования на Python

Лекция 2

Вадим Кохтев
ФКН ВШЭ
Яндекс

Линейная регрессия

- Целевая переменная — вещественная

Линейная регрессия

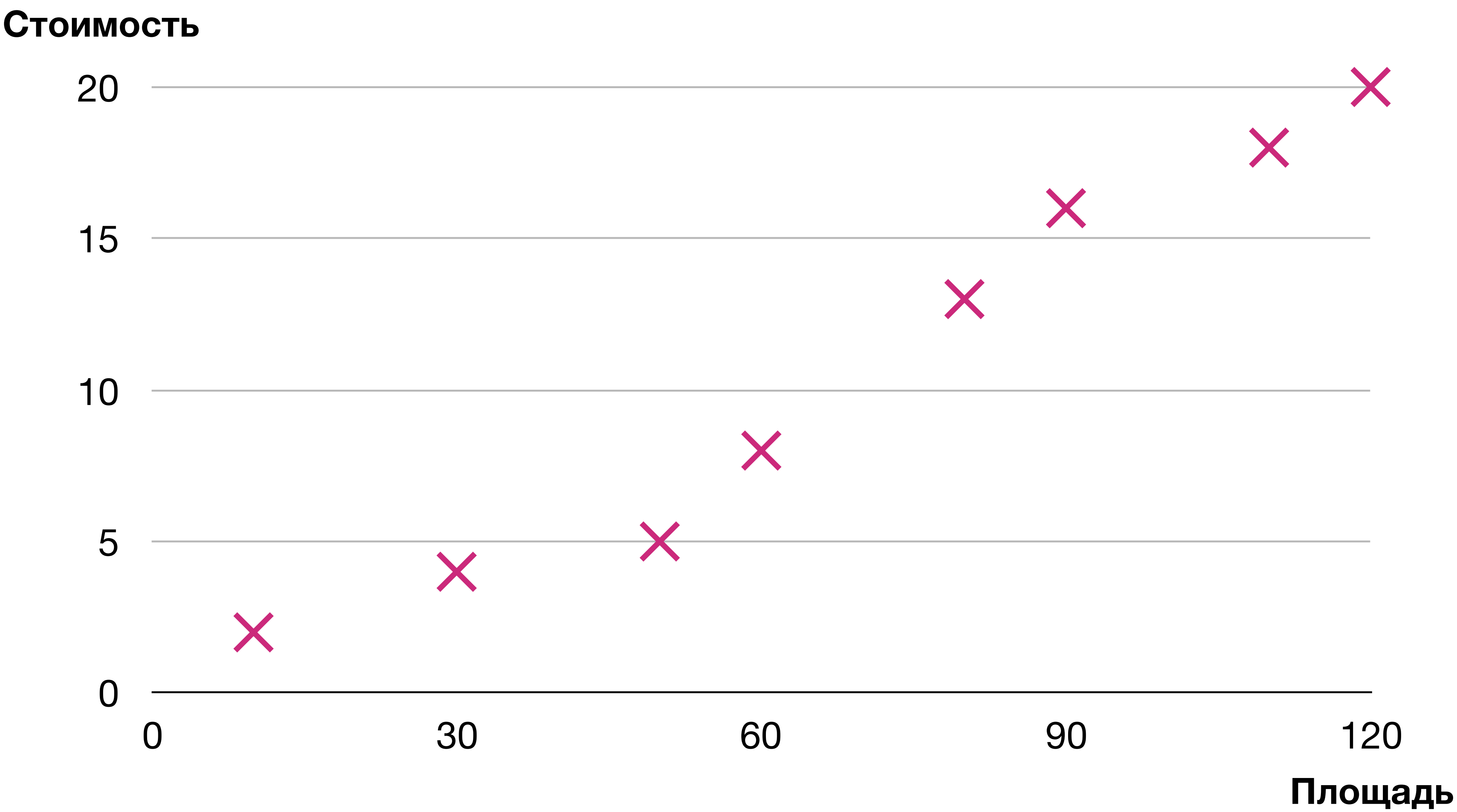
- Целевая переменная — вещественная
- Рассматриваем линейную зависимость от признаков

Линейная регрессия

- Целевая переменная — вещественная
- Рассматриваем линейную зависимость от признаков
- Пример: предсказание стоимости квартиры по площади

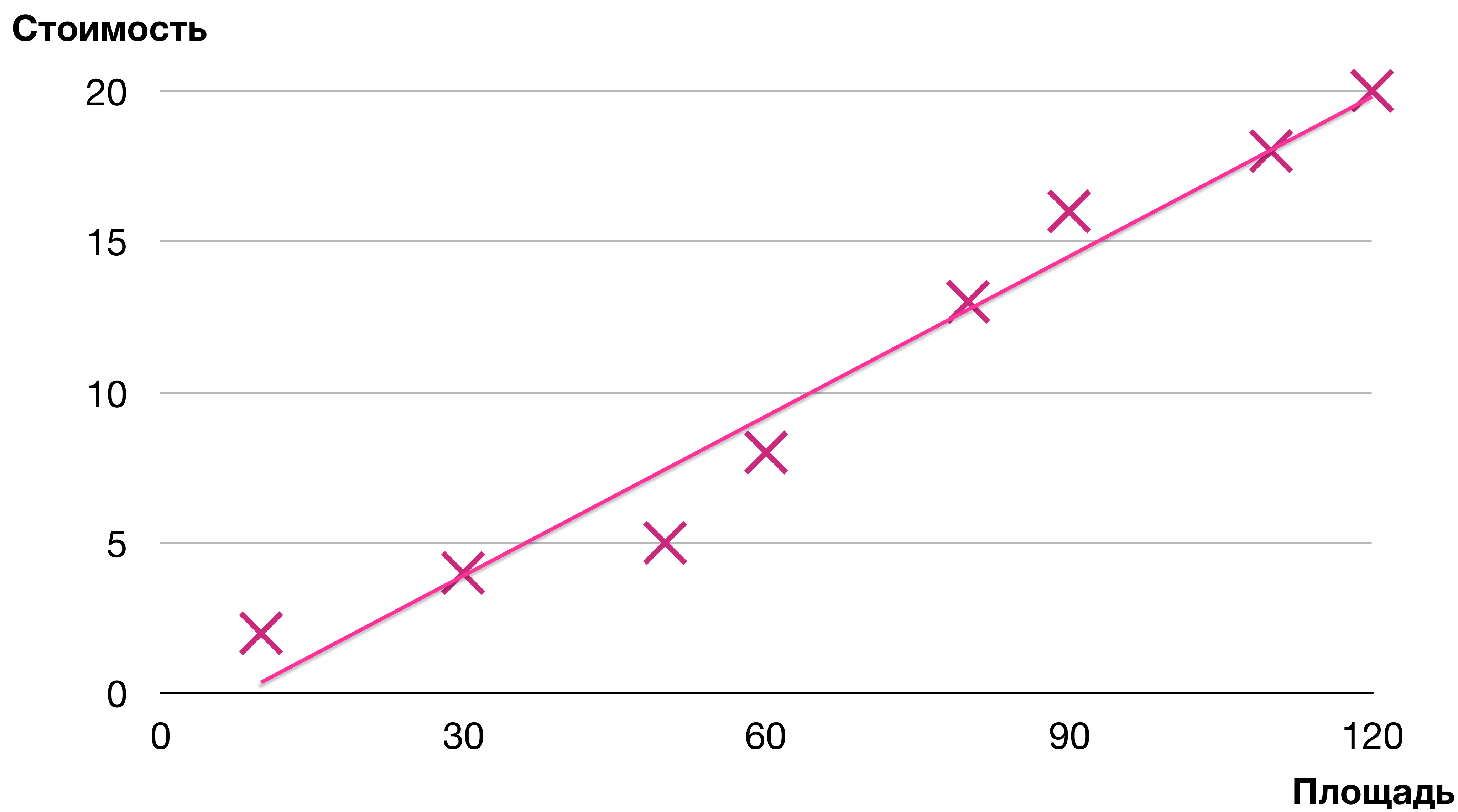
Линейная регрессия

- Целевая переменная — вещественная
- Рассматриваем линейную зависимость от признаков
- Пример: предсказание стоимости квартиры по площади



Линейная регрессия

- Целевая переменная — вещественная
- Рассматриваем линейную зависимость от признаков
- Пример: предсказание стоимости квартиры по площади



Построение линейной регрессии

- А как построить зависимость?

Построение линейной регрессии

- А как построить зависимость?
- Сначала надо понять, когда модель ошибается

Построение линейной регрессии

- А как построить зависимость?
- Сначала надо понять, когда модель ошибается
- Введем функционал ошибки, будем его минимизировать:

$$L(a(x), y) \geq 0$$

Построение линейной регрессии

- А как построить зависимость?
- Сначала надо понять, когда модель ошибается
- Введем функционал ошибки, будем его минимизировать:

$$L(a(x), y) \geq 0$$

$a(x)$ — модель, y — целевая переменная

Построение линейной регрессии

- А как построить зависимость?
- Сначала надо понять, когда модель ошибается
- Введем функционал ошибки, будем его минимизировать:

$$L(a(x), y) \geq 0$$

$a(x)$ — модель, y — целевая переменная

- Можно считать $a(x) - y$, но смысла не будет

Построение линейной регрессии

- А как построить зависимость?
- Сначала надо понять, когда модель ошибается
- Введем функционал ошибки, будем его минимизировать:

$$L(a(x), y) \geq 0$$

$a(x)$ — модель, y — целевая переменная

- Можно считать $a(x) - y$, но смысла не будет
- $|a(x) - y|$ — трудно считать производную

Построение линейной регрессии

- А как построить зависимость?
- Сначала надо понять, когда модель ошибается
- Введем функционал ошибки, будем его минимизировать:

$$L(a(x), y) \geq 0$$

$a(x)$ — модель, y — целевая переменная

- Можно считать $a(x) - y$, но смысла не будет
- $|a(x) - y|$ — трудно считать производную
- Среднеквадратичная ошибка, Mean Squared Error:

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

Метрики качества

- Модель построена, но как оценить качество?

Метрики качества

- Модель построена, но как оценить качество?
- Mean Squared Error (MSE)

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

Метрики качества

- Модель построена, но как оценить качество?
- Mean Squared Error (MSE)

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

	$a(x) = 10$ $y = 20$	$a(x) = 100$ $y = 200$
MSE	100	10000

Метрики качества

- Модель построена, но как оценить качество?
- Mean Squared Error (MSE)

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

- Mean Absolute Error (MAE)

$$L(a(x), y) = \frac{1}{l}\sum_{i=1}^{l}|a(x_i) - y_i|$$

	$a(x) = 10$, $y = 20$	$a(x) = 100$, $y = 200$
MSE	100	10000

Метрики качества

- Модель построена, но как оценить качество?
- Mean Squared Error (MSE)

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

- Mean Absolute Error (MAE)

$$L(a(x), y) = \frac{1}{l}\sum_{i=1}^{l}|a(x_i) - y_i|$$

	$a(x) = 10$ $y = 20$	$a(x) = 100$ $y = 200$
MSE	100	10000
MAE	10	100

Метрики качества

- Модель построена, но как оценить качество?
- Mean Squared Error (MSE)

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

- Mean Absolute Error (MAE)

$$L(a(x), y) = \frac{1}{l}\sum_{i=1}^{l}|a(x_i) - y_i|$$

- Root Mean Squared Error (RMSE)

$$L(a(x), y) = \sqrt{\frac{1}{l}\sum_{i=1}^{l}(a(x_i) - y_i)^2}$$

	$a(x) = 10$ $y = 20$	$a(x) = 100$ $y = 200$
MSE	100	10000
MAE	10	100

Метрики качества

- Модель построена, но как оценить качество?
- Mean Squared Error (MSE)

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

- Mean Absolute Error (MAE)

$$L(a(x), y) = \frac{1}{l} \sum_{i=1}^{l} |\, a(x_i) - y_i \,|$$

- Root Mean Squared Error (RMSE)

$$L(a(x), y) = \sqrt{\frac{1}{l} \sum_{i=1}^{l} (a(x_i) - y_i)^2}$$

	$a(x) = 10$, $y = 20$	$a(x) = 100$, $y = 200$
MSE	100	10000
MAE	10	100
RMSE	3,16	10

Линейная регрессия в матричной форме

- Модель $a(x)$ — произведение весов на признаки

Линейная регрессия в матричной форме

- Модель $a(x)$ — произведение весов на признаки

$$a(x) = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_d x_d$$

Линейная регрессия в матричной форме

- Модель $a(x)$ — произведение весов на признаки

$$a(x) = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_d x_d$$

- $x_1, x_2, \ldots x_d$ — значения признаков

Линейная регрессия в матричной форме

- Модель $a(x)$ — произведение весов на признаки

$$a(x) = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_d x_d$$

- $x_1, x_2, \ldots x_d$ — значения признаков
- $w_1, \ldots + w_d$ — параметры

Линейная регрессия в матричной форме

- Модель $a(x)$ — произведение весов на признаки

$$a(x) = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_d x_d$$

- $x_1, x_2, \ldots x_d$ — значения признаков
- $w_1, \ldots + w_d$ — параметры

- Получается, что $a(x)$ — скалярное произведение!

$$a(x) = < w, x >$$

Линейная регрессия в матричной форме

- Модель $a(x)$ — произведение весов на признаки

$$a(x) = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_d x_d$$

- $x_1, x_2, \ldots x_d$ — значения признаков
- $w_1, \ldots + w_d$ — параметры

- Получается, что $a(x)$ — скалярное произведение!

$$a(x) = < w, x >$$

- При обучении решаем задачу:

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

Линейная регрессия в матричной форме (2)

- Давайте вспомним, как работает матричное умножение

Линейная регрессия в матричной форме (2)

- Давайте вспомним, как работает матричное умножение

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{\ell 1} & x_{\ell 2} & \cdots & x_{\ell d} \end{pmatrix}$$

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix}$$

Линейная регрессия в матричной форме (2)

- Давайте вспомним, как работает матричное умножение

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{\ell 1} & x_{\ell 2} & \cdots & x_{\ell d} \end{pmatrix}$$

$$\begin{pmatrix} \sum_{i=1}^{d} w_i x_{1i} \\ \sum_{i=1}^{d} w_i x_{2i} \\ \vdots \\ \sum_{i=1}^{d} w_i x_{\ell i} \end{pmatrix}$$

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix}$$

Линейная регрессия в матричной форме (2)

- Давайте вспомним, как работает матричное умножение

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{\ell 1} & x_{\ell 2} & \cdots & x_{\ell d} \end{pmatrix}$$

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix}$$

$$\begin{pmatrix} \sum_{i=1}^{d} w_i x_{1i} \\ \sum_{i=1}^{d} w_i x_{2i} \\ \vdots \\ \sum_{i=1}^{d} w_i x_{\ell i} \end{pmatrix} \quad < w, x_1 > = a(x_1)$$

Линейная регрессия в матричной форме (2)

- Давайте вспомним, как работает матричное умножение

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{\ell 1} & x_{\ell 2} & \cdots & x_{\ell d} \end{pmatrix}$$

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix}$$

$$\begin{pmatrix} \sum_{i=1}^{d} w_i x_{1i} \\ \sum_{i=1}^{d} w_i x_{2i} \\ \vdots \\ \sum_{i=1}^{d} w_i x_{\ell i} \end{pmatrix} \quad < w, x_1 > = a(x_1)$$

- Тогда можно записать MSE в матричном виде

$$\frac{1}{l}(Xw - y)^T(Xw - y) \longrightarrow \min_{w}$$

Как минимизировать функции?

Как минимизировать функции?

- Вспомним матанализ

Как минимизировать функции?

- Вспомним матанализ
- Нужно посчитать градиент от функции

Как минимизировать функции?

- Вспомним матанализ
- Нужно посчитать градиент от функции
- Градиент выглядит так:

Как минимизировать функции?

- Вспомним матанализ
- Нужно посчитать градиент от функции
- Градиент выглядит так:

$$\nabla Q(X, w) = (\frac{\partial Q}{\partial w_1}, \frac{\partial Q}{\partial w_2}, \ldots \frac{\partial Q}{\partial w_d})$$

Как минимизировать функции?

- Вспомним матанализ
- Нужно посчитать градиент от функции
- Градиент выглядит так:

$$\nabla Q(X, w) = (\frac{\partial Q}{\partial w_1}, \frac{\partial Q}{\partial w_2}, \ldots \frac{\partial Q}{\partial w_d})$$

- Минимум $\nabla Q(X, w) = 0$

Как минимизировать функции?

- Вспомним матанализ
- Нужно посчитать градиент от функции
- Градиент выглядит так:

$$\nabla Q(X, w) = (\frac{\partial Q}{\partial w_1}, \frac{\partial Q}{\partial w_2}, \ldots \frac{\partial Q}{\partial w_d})$$

- Минимум $\nabla Q(X, w) = 0$
- Если проделать такое с линейной регрессией, получим:

Как минимизировать функции?

- Вспомним матанализ
- Нужно посчитать градиент от функции
- Градиент выглядит так:

$$\nabla Q(X, w) = (\frac{\partial Q}{\partial w_1}, \frac{\partial Q}{\partial w_2}, \ldots \frac{\partial Q}{\partial w_d})$$

- Минимум $\nabla Q(X, w) = 0$
- Если проделать такое с линейной регрессией, получим:

$$\nabla Q = \frac{2}{l} X^T (Xw - y) = 0$$

Как минимизировать функции?

- Вспомним матанализ
- Нужно посчитать градиент от функции
- Градиент выглядит так:

$$\nabla Q(X, w) = (\frac{\partial Q}{\partial w_1}, \frac{\partial Q}{\partial w_2}, \ldots \frac{\partial Q}{\partial w_d})$$

- Минимум $\nabla Q(X, w) = 0$
- Если проделать такое с линейной регрессией, получим:

$$\nabla Q = \frac{2}{l} X^T (Xw - y) = 0$$

- Можно получить формулу для весов!

Как минимизировать функции?

- Вспомним матанализ
- Нужно посчитать градиент от функции
- Градиент выглядит так:

$$\nabla Q(X, w) = (\frac{\partial Q}{\partial w_1}, \frac{\partial Q}{\partial w_2}, \ldots \frac{\partial Q}{\partial w_d})$$

- Минимум $\nabla Q(X, w) = 0$
- Если проделать такое с линейной регрессией, получим:

$$\nabla Q = \frac{2}{l} X^T (Xw - y) = 0$$

- Можно получить формулу для весов!

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

Резюме

- Линейная регрессия — простая модель машинного обучения
- Обсудили метрики качества
- Вспомнили математику

Организационное

http://wiki.cs.hse.ru/Основы_программирования_в_Python_2019
https://t.me/PythonWE2019