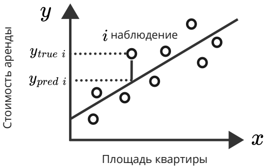
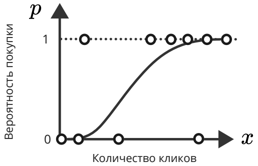
Занятие 2	Занятие 3
Линейная регрессия	**Логистическая регрессия**





Форма датасета

n	Площадь квартиры (x)	Реальная стоимость аренды (y_true)	Предсказанная стоимость аренды (y_pred)
1	50	60 000	58 654
2	28	25 000	23 634
...	...	...	...
N	35	28 000	31 634

n	Количество кликов (x)	Покупка (t)	Предсказанная вероятность (p)
1	50	1	0.85
2	5	0	0.28
...	...	...	...
N	67	1	0.64

Функция для минимизации

$$Loss(w_0, w_1) = \sum_{i=1}^{N} \left(y_{true\ i} - y_{pred\ i}\right)^2$$

$$y_{pred\ i} = w_0 + w_1 \cdot x$$

Линейная модель предсказывает целевую переменную

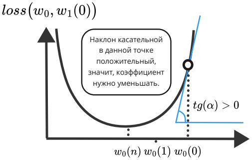
$$Loss(w_0, w_1) = -\sum_{i=1}^{N} (t_i \cdot ln(p_i) + \\ +(1 - t_i) \cdot ln(1 - p_i))$$

$$ln\left(\frac{p_i}{1 - p_i}\right) = w_0 + w_1 \cdot x$$

Линейная модель предсказывает логарифм шансов быть объектом класса 1

Подбор коэффициентов осуществляется методом градиентного спуска.

Градиентный спуск - численный метод нахождения минимума функции



1. Сперва коэффициенты w_0 и w_1 инициализируются случайным образом:

$$w_0 = w_0(0) \qquad w_1 = w_1(0)$$

2. Считается тангенс наклона касательной вдоль направлений w_0 и w_1
3. Происходит обновление w_0 и w_1

$$w_0(1) = w_0(0) - \underbrace{\boxed{\alpha}}_{\text{learning rate}} \cdot \underbrace{\boxed{\left.\frac{\partial\, loss(w_0, w_1)}{\partial w_0}\right|_{[w_0(0), w_1(0)]}}}_{tg(\alpha)}$$

(коэффициент скорости обучения)

$$w_1(1) = w_1(0) - \alpha \left.\frac{\partial\, loss(w_0, w_1)}{\partial w_1}\right|_{w_0(0), w_1(0)}$$

Чтобы лучше понять градиентный спуск, посмотри второе видео в плейлисте, посвящённом обучению нейронных сетей, от 3Blue1Brow. Длительность видео 21 минута.

Занятие 2	Занятие 3

Метрики регрессии

Средняя абсолютная ошибка

$$MAE = \sum_{i=1}^{N} \left| y_{true\ i} - y_{pred\ i} \right|$$

Коэффициент множественной корреляции

$$R = corr(y_{true}, y_{pred})$$

Коэффициент детерминации - квадрат коэффициента множественной корреляции. Можно показать, что:

$$\boxed{R^2 = 1 - \frac{\sum_{i=1}^{N} \left(y_{true\ i} - y_{pred\ i}\right)^2}{\sum_{i=1}^{N} \left(y_{true\ i} - \bar{y}\right)^2}}$$

Показывает, насколько модель работает лучше, чем предсказание среднего.

Если для каждого примера предсказывать среднее, то R^2 = 0. Если делать хорошие предсказания, то R^2 --> 1.

Посмотри это 20-минутное видео от 3Blue1Brown, чтобы лучше понять метрики классификации.

Метрики классификации

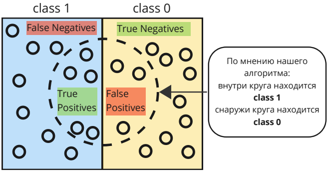
ROC AUC показывает вероятность того, что если взять один объект класса 0 и один объект класса 1, то score на классе 1 будет больше скора на класса 0.

ROC AUC Random = 0.5

ROC AUC Ideal = 1.0

Метрики для меток.

Пусть при определённом пороге принятия решения мы получили картинку:



Фундаментальные метрики классификатора

$$TPR = \frac{TP}{TP + FN} \quad (Sensitivity\ or\ Recall)$$

$$TNR = \frac{TN}{TN + FP} \quad (Specifity)$$

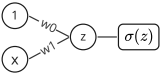
Любые другие метрики классификации выражаются через TPR, TNR и Prevalence

$$Precision = \frac{TP}{TP + FP} \approx \underbrace{\boxed{Prevalence}}_{\text{Распространённость целевого класса}} \cdot \frac{TPR}{1 - TNR}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Мульти-класс классификация

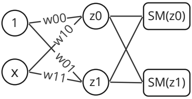
Логистическая регрессия (на 2 класса)



$$z = w_0 + w_1 \cdot x$$

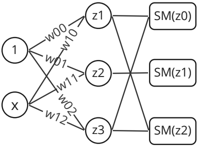
$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Мульти-класс логистическая регрессия на 2 класса



$$SM(z_i) = \frac{e^{z_i}}{e^{z_1} + e^{z_2}}$$

Мульти-класс логистическая регрессия на 3 класса



❗ Можно показать, что эти 2 подхода идентичны ❗