Задача классификации

2021

Что будет ...

Задача классификации

- **Логистическая регрессия**
- **Метрики**

Классификация

Обучающая выборка - множество объектов, разделенные на классы

Задача - определить класс нового объекта

Бинарная классификация - класса задано только 2

Многоклассовая классификация - классов более 2

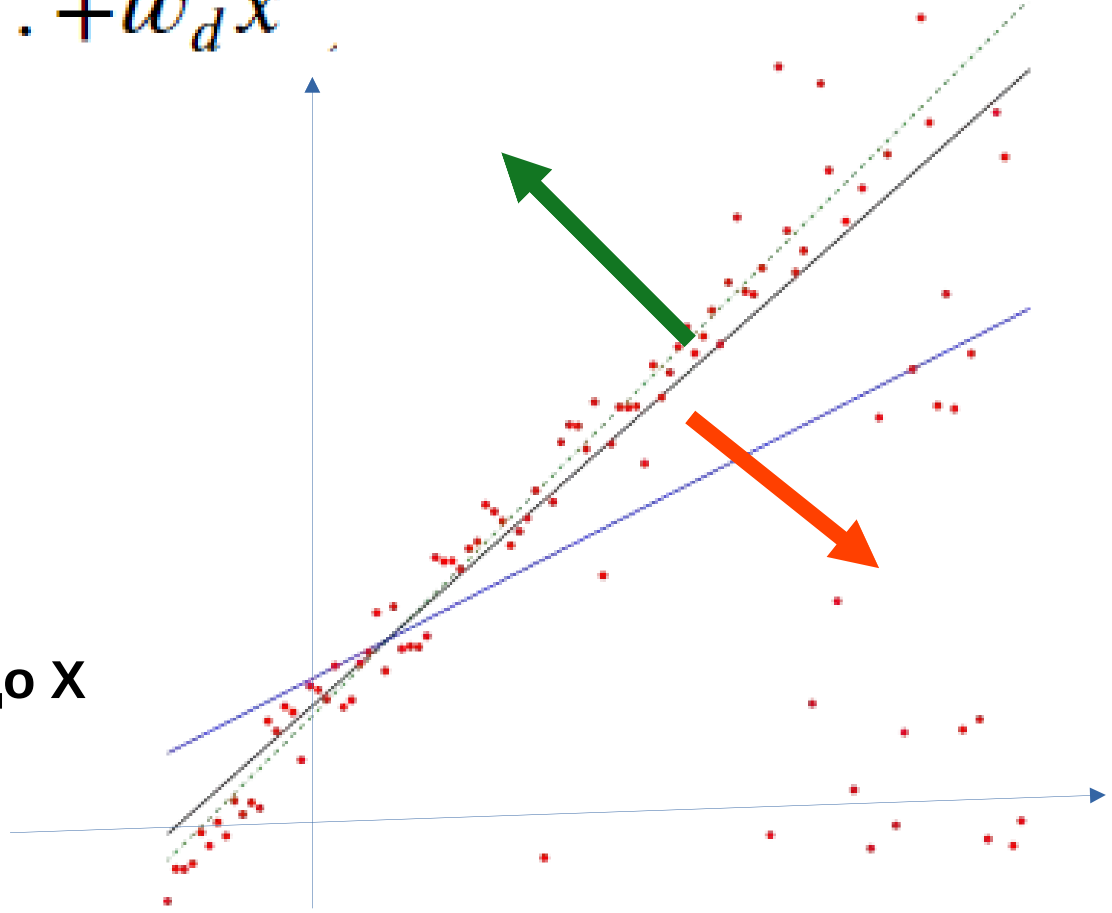
Линейная модель

$$a(x) = w_0 + w_1 x^1 + w_2 x^2 + \ldots + w_d x^d$$

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

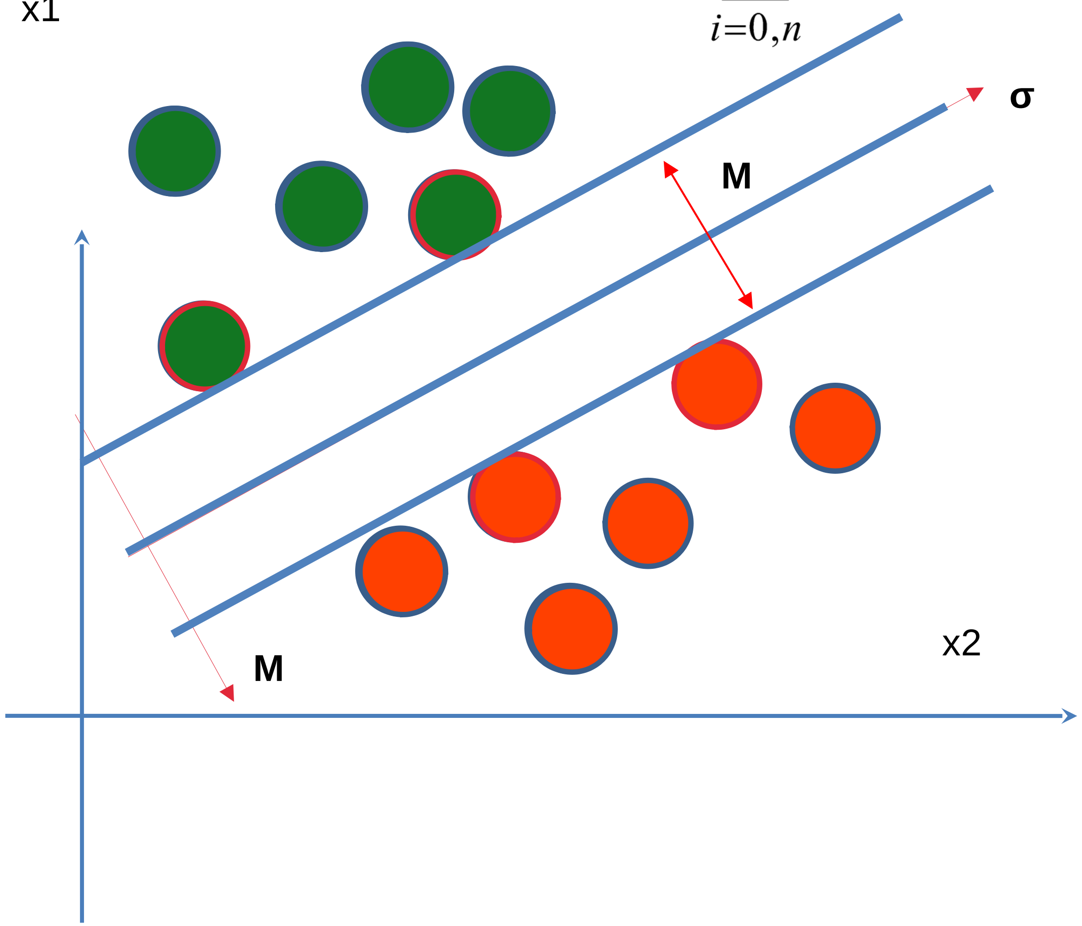
a(x) = 0 гиперплоскость

$\dfrac{|a(X)|}{||w||}$ **расстояние от гиперплоскости до X**



Отступ (Зазор - Margin)

$$y = \sum_{i=0,n} X_i w_i = 0$$

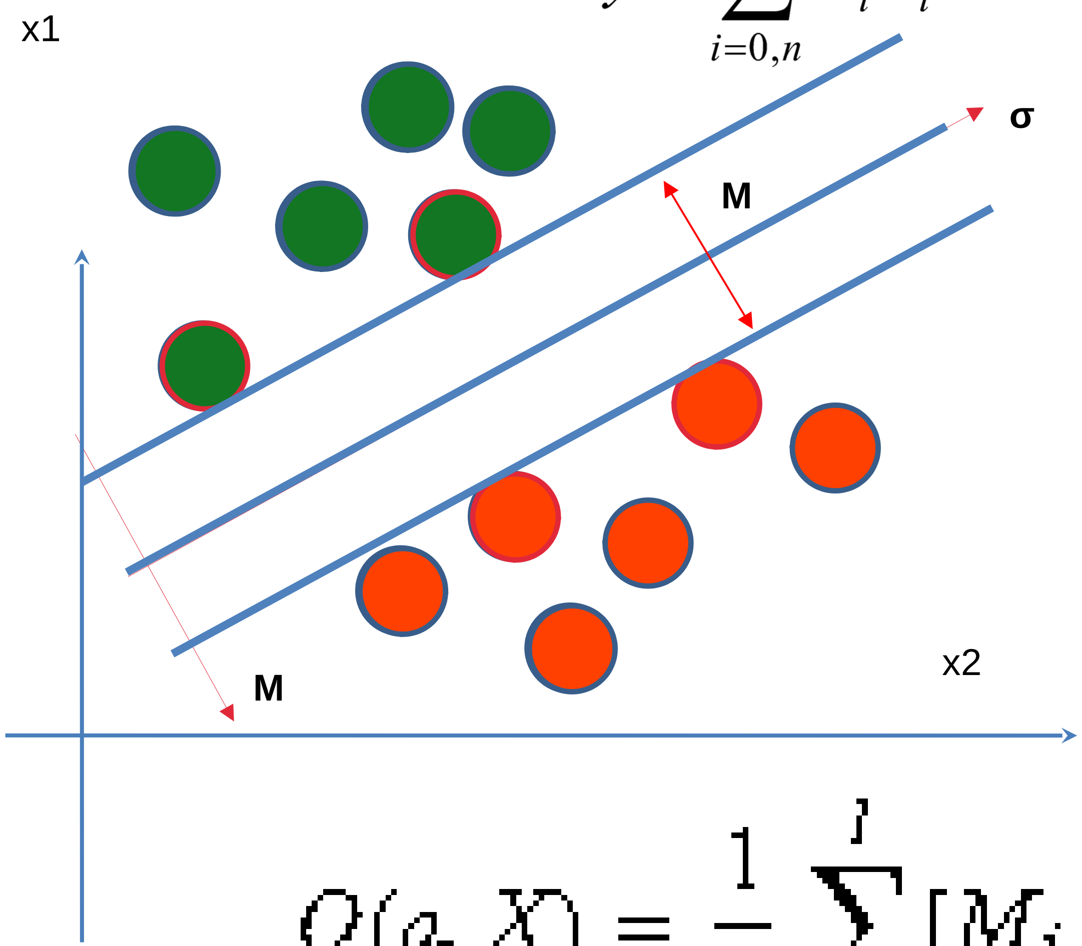


$$y_i : \{-1,1\}$$

$$M_i = y_i \cdot \langle X_i \cdot W \rangle$$

Классификация - Потери

$$y = \sum_{i=0,n} X_i w_i = 0$$



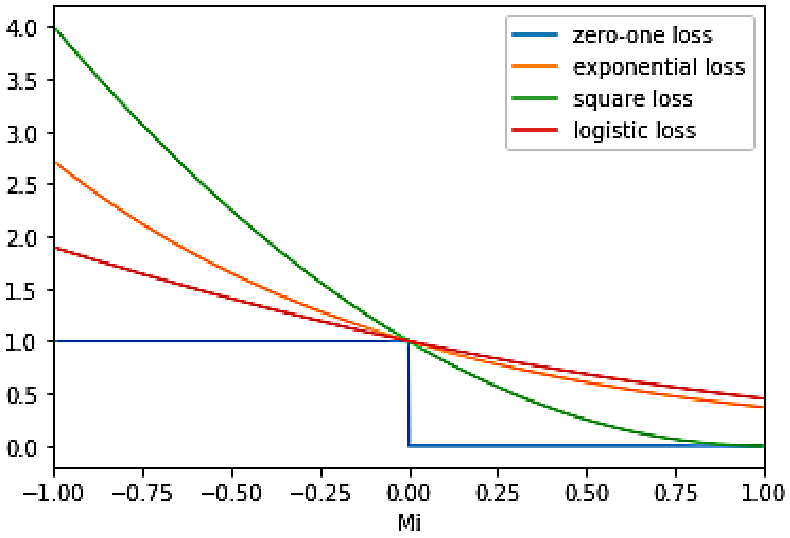
$$Q(a, X) = \frac{1}{l}\sum_{i=1}^{l}[a(x_i) \neq y_i]$$

$$y_i: \{-1,1\}$$

$$M_i = y_i \cdot \langle X_i \cdot W \rangle$$

$$Q(a, X) = \frac{1}{l}\sum_{i=1}^{l}[M_i < 0] = \frac{1}{l}\sum_{i=1}^{l}[y_i \langle w, x_i \rangle < 0].$$

Рабочие Потери



$$[M_i < 0] \leq \breve{L}(M_i).$$

$$Q(a, X) \leq \breve{Q}(a, X) = \frac{1}{l}\sum_{i=1}^{l} \breve{L}(M_i) \to \min_{\omega}$$

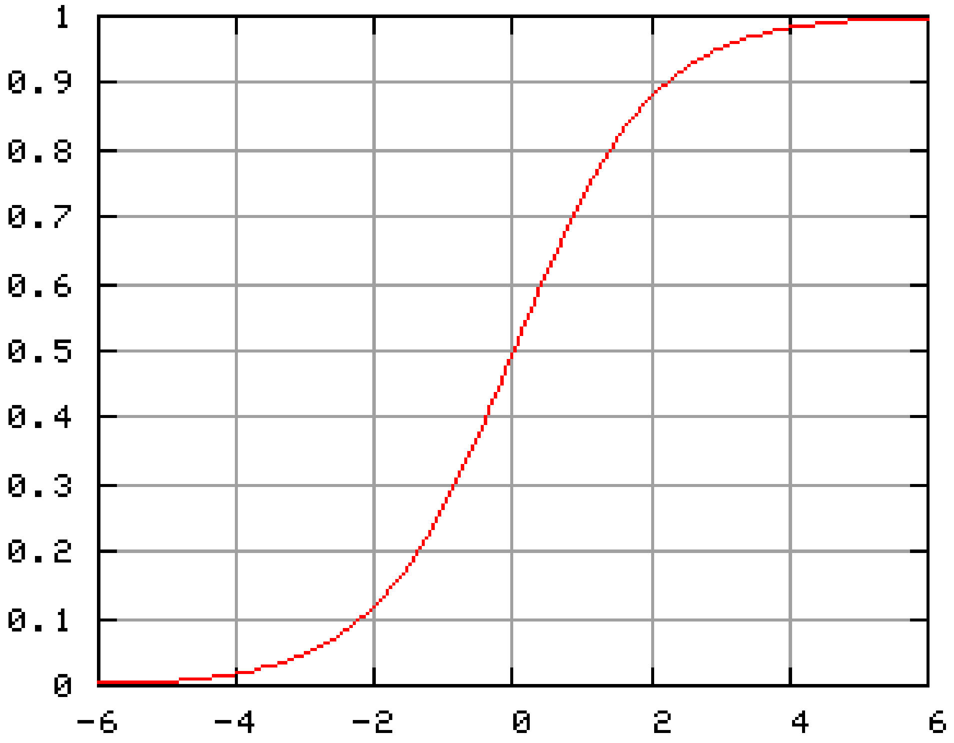
- *экспоненциальная функция потерь* $\breve{L}(M_i) = \exp(-M_i)$
- *квадратичная функция потерь* $\breve{L}(M_i) = (1 - (M_i))^2$
- *логистическая функция потерь* $\breve{L}(M_i) = \log_2(1 + \exp(-M_i))$

Логистическая регрессия

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

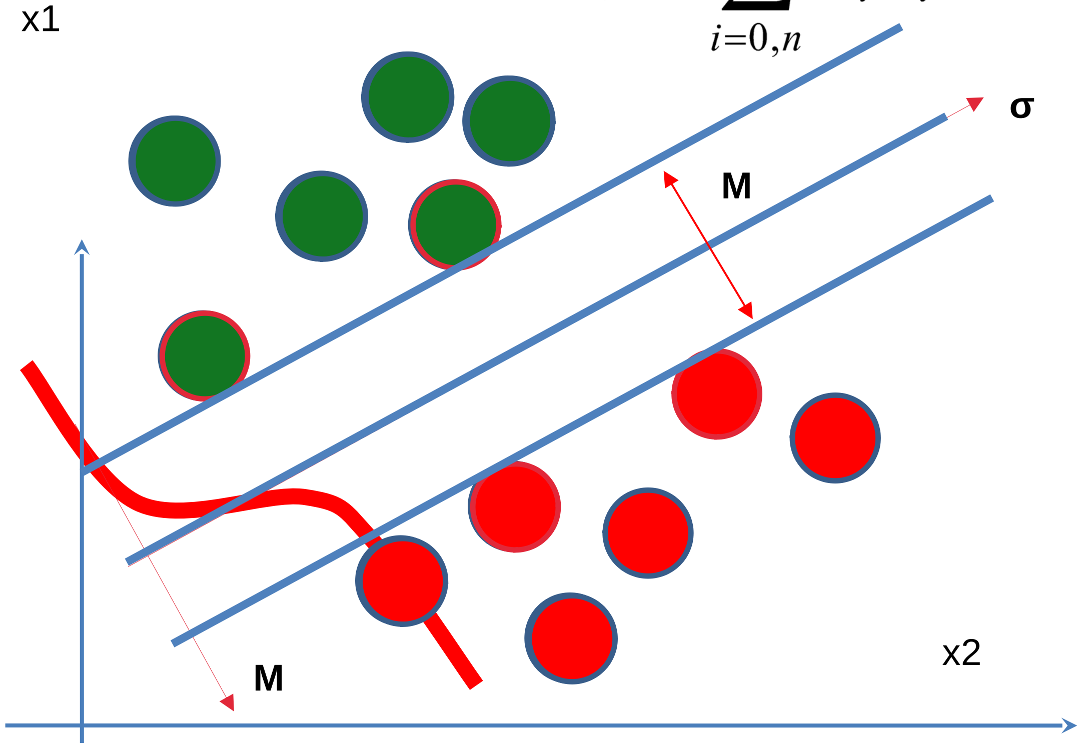
Наша модель

$$a = sigmoid(w^T x) = \frac{1}{1+e^{-w^T x}}$$



Логистическая регрессии

$$y = \sum_{i=0,n} X_i w_i = 0$$



$$\sigma = \frac{1}{1+e^{-M_i}}$$

$$M_i = y_i \cdot \langle X_i \cdot W \rangle$$

$$p_+ = \sigma(\langle w, x_i \rangle) = \frac{1}{1 + exp(-\langle w, x_i \rangle)}.$$

$$\langle w, x_i \rangle = \ln \frac{p_+}{1 - p_+}$$

Log Loss

Метод максимального правдоподобия

Рассматриваем задачу классификации с двумя классами:-1 и 1.

Выборка (по Бернулли): каждый объект с вероятностью p_i относится к классу 1 и с вероятностью $(1–p_i)$ к классу -1.

Log Loss

Метод максимального правдоподобия

Результат моде $a_i = a(x_i|w)$ **где w - параметры модели**

Функция правдоподоб $P(y = y_i|x_i) = p_+^{[y_i=+1]}(1 - p_+)^{[y_i=-1]}$

$$P(y|X) = L(X) = \prod_{i=1}^{l} p_+^{[y_i=+1]}(1 - p_+)^{[y_i=-1]}.$$

$$p(y|X,w) = \prod_i p(y_i|x_i,w) = \prod_i a_i^{y_i}(1 - a_i)^{1-y_i}$$

Для конкретной a(xi|w)
с классами по yi = {0,1}

Log Loss

Метод максимального правдоподобия

Метод максимального правдоподобия:

$$-\ln L(X) = -\sum_{i=1}^{l}([y_i = +1]\ln p_+ + [y_i = -1]\ln(1 - p_+)).$$

$$-\ln L(X) = -\sum_{i=1}^{l}([y_i = +1]\ln\frac{1}{1 + exp(-\langle w, x_i\rangle)} + [y_i = -1]\ln(1 - \frac{1}{1 + exp(-\langle w, x_i\rangle)})) =$$

$$= \sum_{i=1}^{l}\ln(1 + exp(-y_i\langle w, x_i\rangle))$$

Log Loss

Метод максимального правдоподобия

Метод максимального правдоподобия:

$$-\ln L(X) = -\sum_{i=1}^{l}([y_i = +1]\ln\frac{1}{1+exp(-\langle w, x_i\rangle)} + [y_i = -1]\ln(1 - \frac{1}{1+exp(-\langle w, x_i\rangle)})) =$$

$$= \sum_{i=1}^{l}\ln(1 + exp(-y_i\langle w, x_i\rangle))$$

$$[M < 0] \le \log_2\left(1 + e^{-M}\right)$$

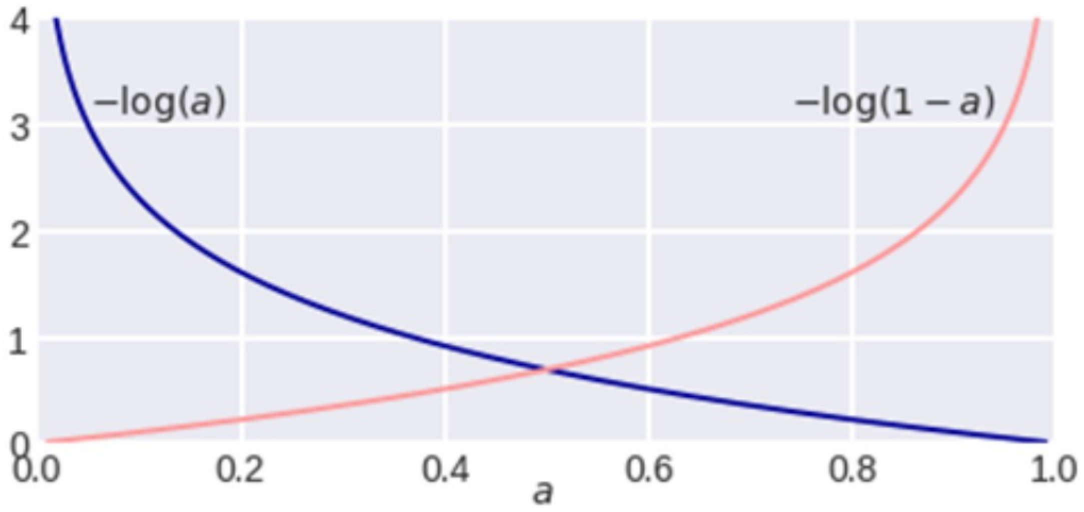
Функции равны до коэффициента 1/ln(2)

Log Loss

Метод максимального правдоподобия

$$-\ln L(X) = -\sum_{i=1}^{l}\left(y_i \ln\frac{1}{1+exp(-\langle w, x_i\rangle)} + (1-y_i)\ln\frac{exp(-\langle w, x_i\rangle)}{1+exp(-\langle w, x_i\rangle)}\right)$$

$$\mathrm{LogLoss} = \sum_i -y_i \log a_i - (1-y_i)\log(1-a_i) = -\begin{cases} \log a_i, & y_i = 1 \\ \log(1-a_i), & y_i = 0 \end{cases}$$



Log Loss

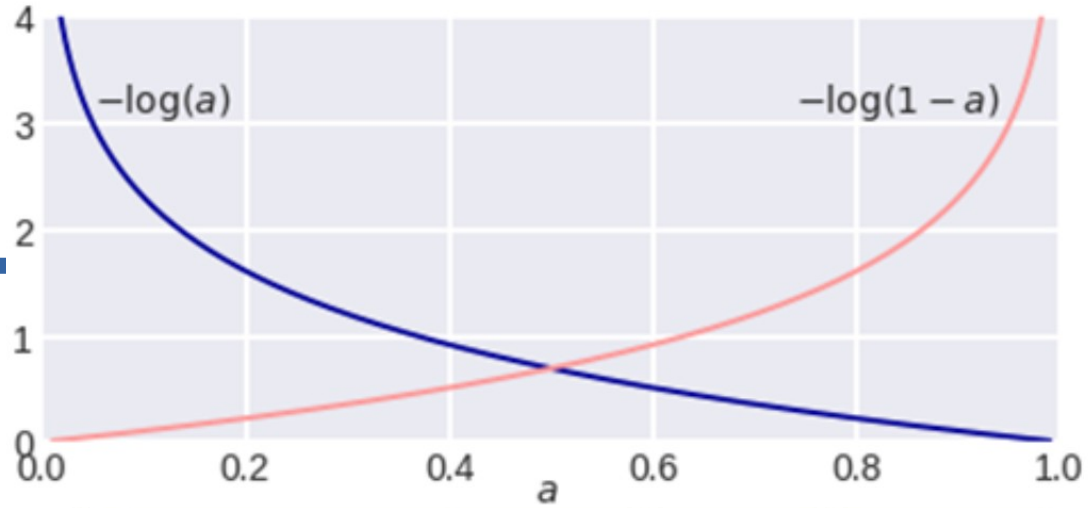
Метод максимального правдоподобия

$$p = y$$

$$-p\log(a_i) - (1-p)\log(1-a_i)$$

$$\frac{p}{a_i} - \frac{1-p}{1-a_i} = 0$$

$$a_i = p$$



Log Loss

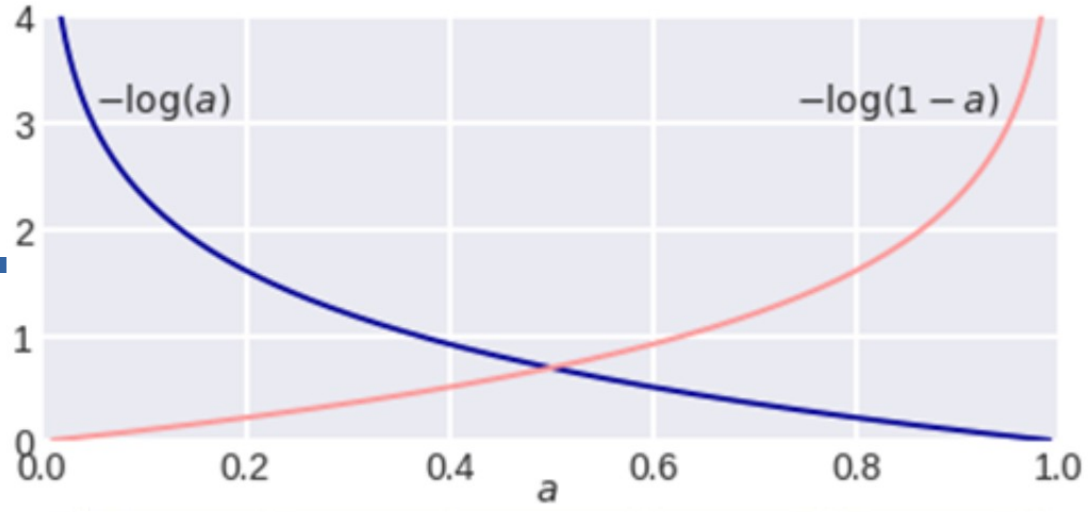
Метод максимального правдоподобия

$$a = p$$

Энтропия

$$-p\log(p)-(1-p)\log(1-p).$$

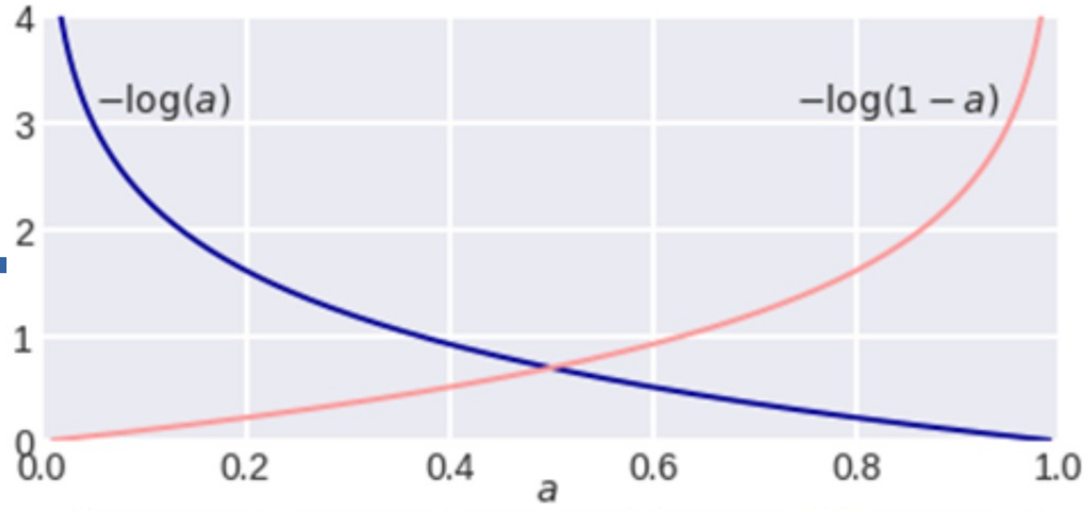
$$H(y,a) = -\sum_x y(x)\log a(x,w)$$



Log Loss

Метод максимального правдоподобия

$$\text{logloss} = -\frac{1}{q}\sum_{i=1}^{q}\sum_{j=1}^{l} y_{ij}\log a_{ij}$$



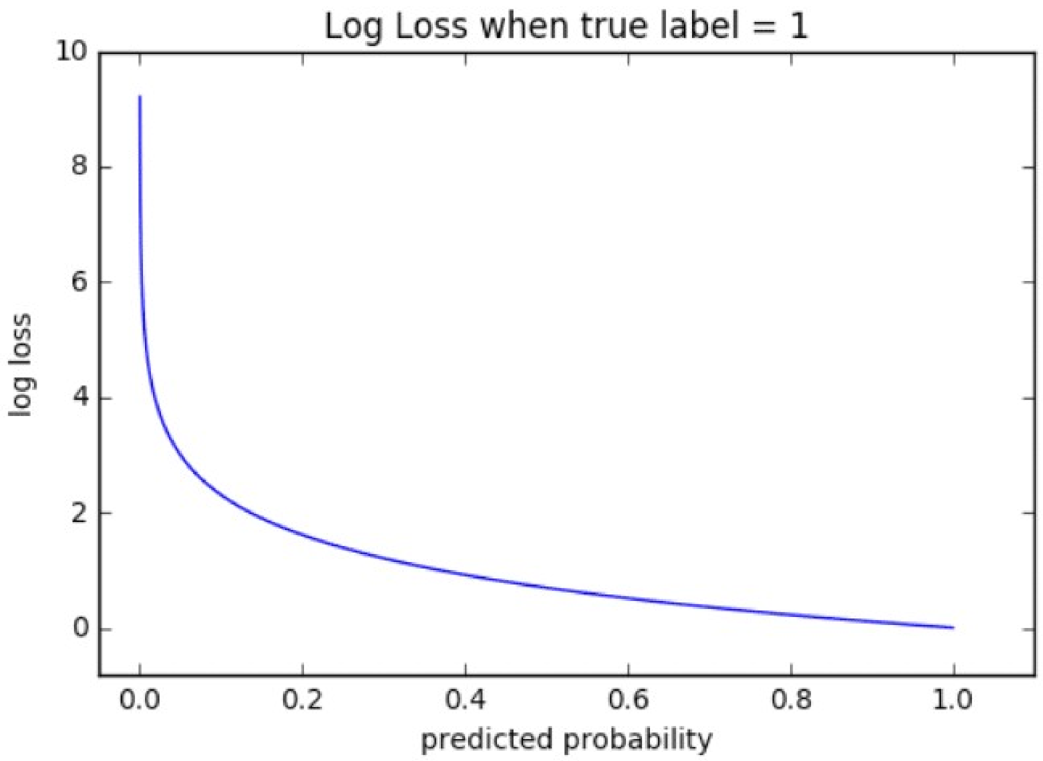
Градиентный спуск

LogLoss - функкнция ошибок

Функцию ошибок надо минимизиро

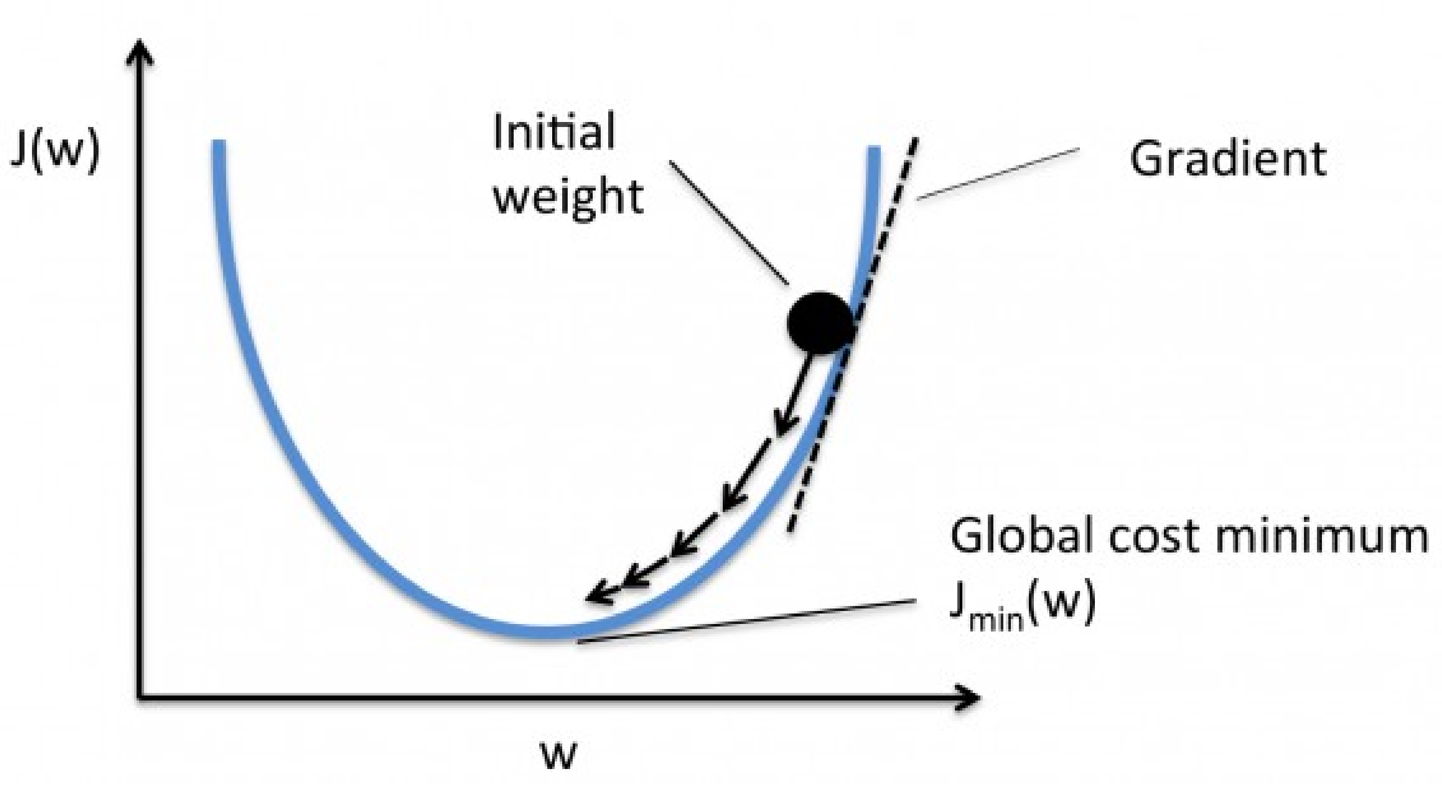
$$w_{n+1} = w_n - \eta \frac{1}{l} X(A - Y)^T$$

$$A = \frac{1}{1+exp(-\langle w, x_i \rangle)},$$



Градиентный спуск

Градиент (производная) показывает направление роста функции.



Оптимизация логистической регрессии

Функция потерь: $$\mathrm{LogLoss} = \sum_i -y_i \log a_i - (1 - y_i) \log(1 - a_i)$$

Логистическая регрессия: $$a = \frac{1}{1 + e^{-w^T x}}$$

Градиент: $$\frac{\partial LogLoss}{\partial w} = (a - y)x$$

Метрики классификации

	Истина +	Истина -
Предсказано +	True positive	False positive
Предсказано -	False negative	True negative

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

$$Precision = Sensitivity = \frac{TP}{TP + FP}$$

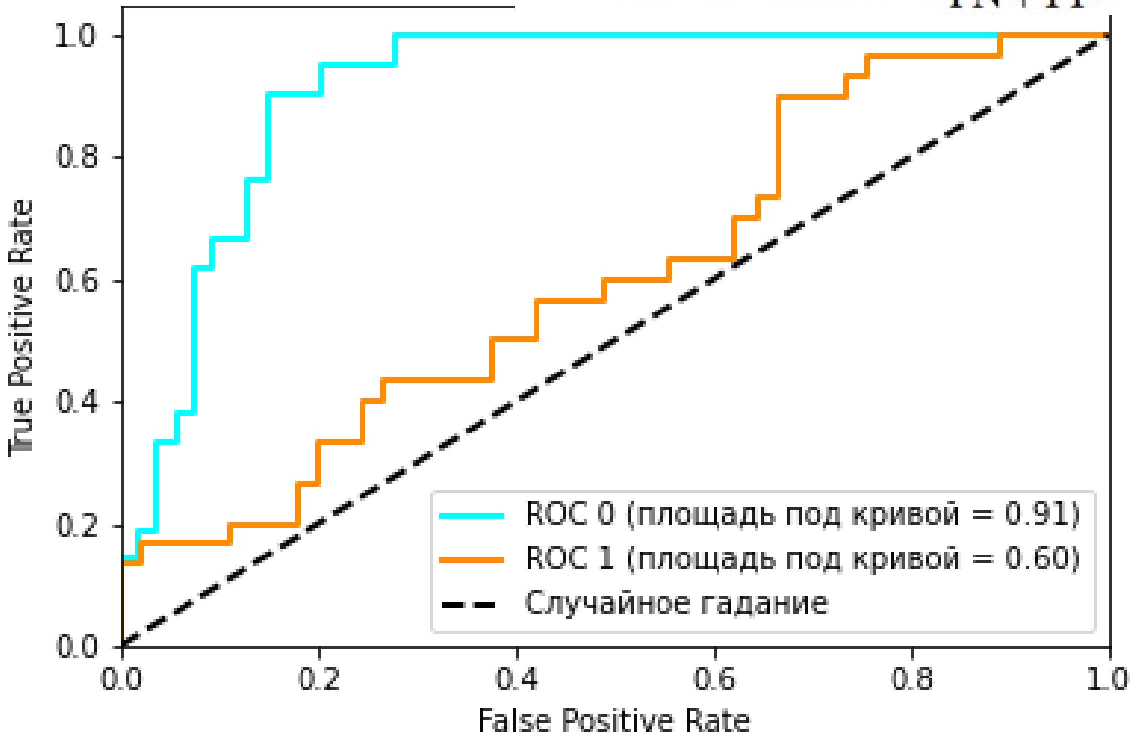
$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

Метрики классификации

$$\text{TPR (sensitivity)} = \frac{TP}{TP + FN}$$

$$\text{FPR (1-specificity)} = \frac{FP}{TN + FP}$$



ROC кривая - зависимость верно классифицируемых объектов положительного класса (Sensitivity) от ложноположительно классифицируемых объектов негативного класса (Specificity)

AUC (Area Under Curve) - площадь под кривой

ROC-AUC - площадь под ROC кривой, численная оценка ROC метрики

Логистическая регрессия

- **Линейный алгоритм классификации оказывается оптимальным байесовским классификатором.**
- **Однозначно определяется вид функции активации (сигмоидная функция) и функции потерь.**
- **Делает численные оценки вероятности его принадлежности каждому из классов.**
- **Является частным случаем обобщённой линейной модели регрессии.**

Логистическая регрессия

- Оценки вероятностей и рисков могут оказаться неадекватными, если признаки зависимы
- Все недостатки метода стохастического градиента.
- Практичная реализация должна предусматривать :
  - стандартизацию данных,
  - отсев выбросов,
  - регуляризацию,
  - отбор признаков.

Логистическая регрессия

ВОПРОСЫ?

Спасибо!

GeekBrains