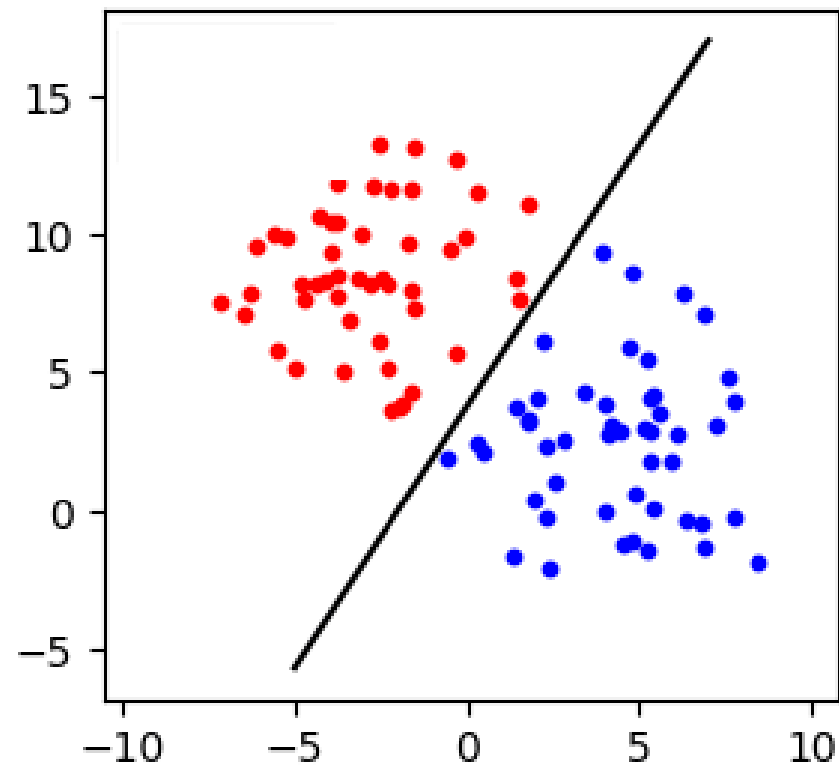




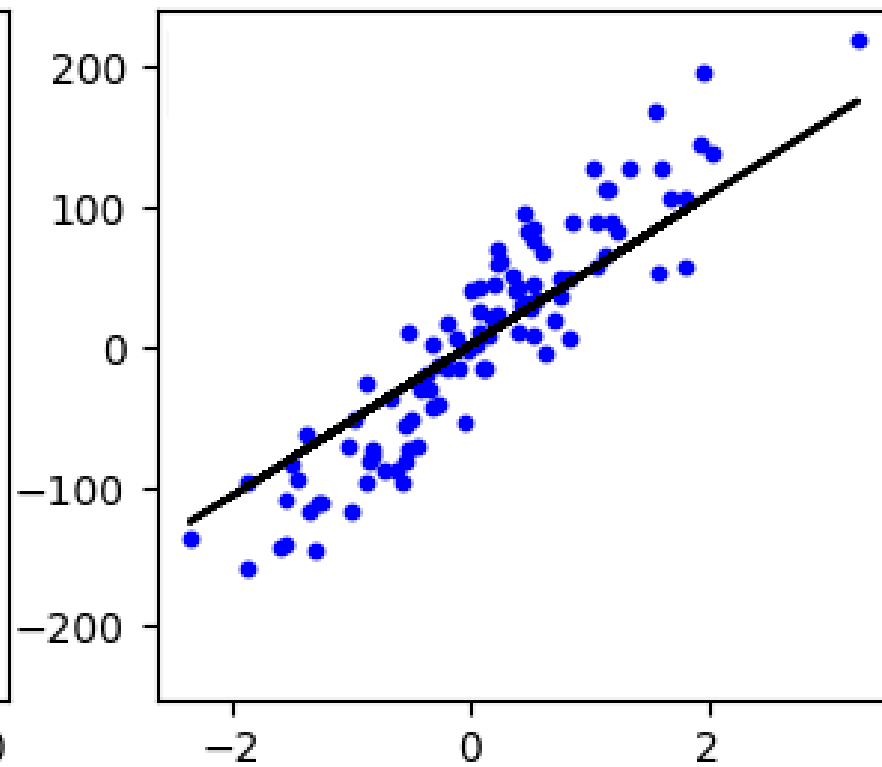
# Интеллектуальные информационные системы

Линейные модели классификации  
Байесовский классификатор

Classification



Regression



# Линейная модель классификации

Пусть  $X = \mathbb{R}^D$  - пространство объектов

$Y = \{-1, +1\}$  - множество допустимых ответов

$X = \{(x_i, y_i)\}_{i=1}^l$  - обучающая выборка


Линейная модель классификации:

$$a(x) = \textit{sign}(\langle w, x \rangle + w_0) = \textit{sign}\left(\sum_{j=1}^D w_j x_j + w_0\right)$$

$$a(x) = \textit{sign}\langle w, x \rangle$$

# Обучение линейного классификатора

- Функционал качества

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


Такой функционал хочется максимизировать

Удобнее решать задачу минимизации, поэтому:



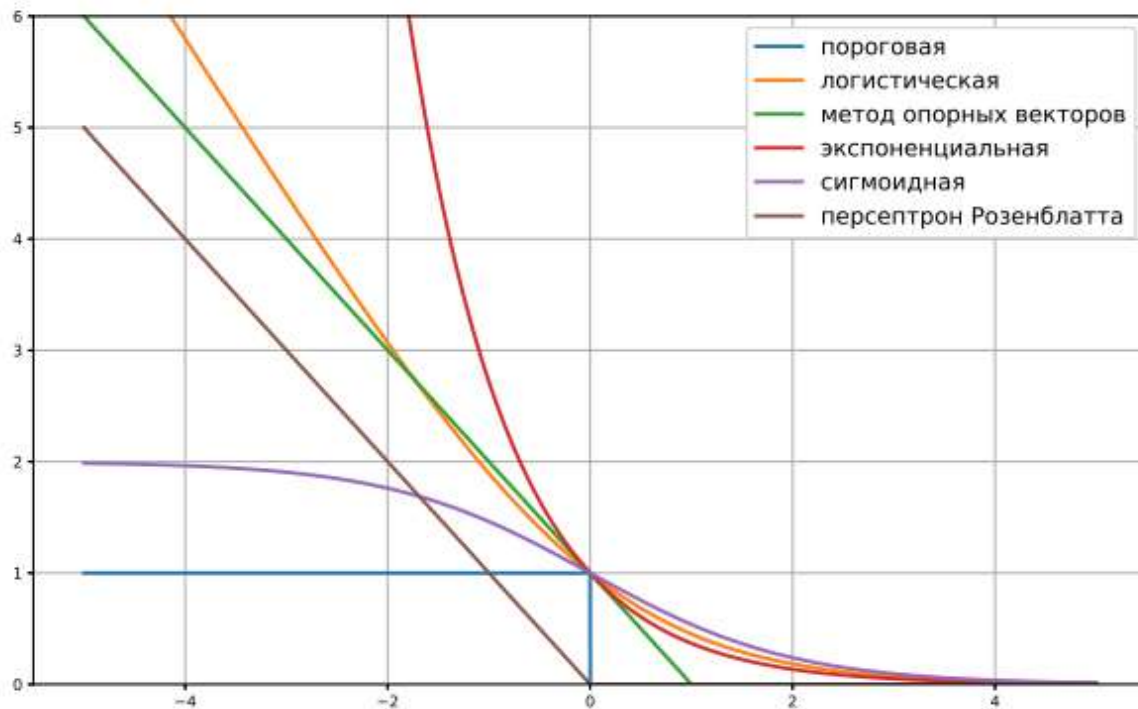
$$Q(a, X) = \frac{1}{l} \sum_{i=1}^l [a(x_i) \neq y_i] = \frac{1}{l} \sum_{i=1}^l [\text{sign}\langle w, x_i \rangle \neq y_i] \rightarrow \min_w$$

- Концепт «margin» - отступ:

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

$M_i = y_i \langle w, x_i \rangle$  - отступ –  
«уверенность» классификатора

# Непрерывные аппроксимации пороговой функции потерь



1.  $[M < 0]$  - пороговая
2.  $L(M) = \log(1 + e^{-M})$  - логистическая
3.  $L(M) = \max(0, 1 - M)$  – SVM
4.  $L(M) = e^{-M}$  - экспоненциальная
5.  $L(M) = 2/(1+e^M)$  – сигмоидная
6.  $L(M) = \max(0, -M)$  – персептрон Розенблатта

# Логистическая регрессия

- Logit – логарифм отношения вероятности положительного события к отрицательному

$$\log \left( \frac{p}{1-p} \right) \longleftarrow \text{ответ модели}$$

$$\langle w, x_i \rangle = \log \left( \frac{p}{1-p} \right) \Rightarrow e^{\langle w, x_i \rangle} = \frac{p}{1-p} \Rightarrow p = \frac{1}{1+e^{-\langle w, x_i \rangle}}$$

$$\boxed{\sigma(z) = \frac{1}{1+e^{-z}}} \longleftarrow \text{сигмоида}$$

$$p = \sigma(\langle w, x_i \rangle)$$

Правдоподобие – позволяет понять, насколько вероятно получить данные для значения целевой переменной  $y$  при данных  $X$  и весах  $w$ .

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

$p_i$  - вероятность, посчитанная из ответов модели

$$l(w, X, y) = \sum_i (y_i \log(\sigma(\langle w, x_i \rangle)) + (1 - y_i) \log(\sigma(-\langle w, x_i \rangle)))$$

Логарифмическое правдоподобие, хотим чтобы было максимальное

Умножим на -1 и будем минимизировать с помощью градиентного спуска

$$L(w, X, y) = - \sum_i (y_i \log(\sigma(\langle w, x_i \rangle)) + (1 - y_i) \log(\sigma(-\langle w, x_i \rangle)))$$

Логистическая функция потерь (log-loss)



# Многоклассовая классификация

## Сведение к набору бинарных задач:

1. Один против всех (one-vs-all / one-vs-rest):

К линейным классификаторов  $b_1(x), \dots, b_k(x)$  вида:

$$b_k(x) = \textit{sign}(\langle w_k, x \rangle + w_{0k})$$

Выбираем самый «уверенный»:

$$a(x) = \textit{argmax}_k (\langle w_k, x \rangle + w_{0k})$$

2. Все против всех (all-vs-all / one-vs-one):

$C_K^2$  линейных классификаторов  $a_{ij}(x), i, j = 1, \dots, K, i \neq j$ .

Выбираем по количеству голосов:

$$a(x) = \textit{argmax}_{k \in \{1, \dots, K\}} \sum_{i=1}^K \sum_{j \neq i} [a_{ij}(x) = k]$$

# Многоклассовая логистическая регрессия

Функция Softmax:

$$\text{Softmax}(z_1, \dots, z_K) = \left( \frac{\exp(z_1)}{\sum_{k=1}^K \exp(z_k)}, \dots, \frac{\exp(z_K)}{\sum_{k=1}^K \exp(z_k)} \right)$$

Вероятность ***k-go*** класса:

$$P(y = k|x, w) = \frac{\exp(\langle w_k, x \rangle + w_{0k})}{\sum_{j=1}^K \exp(\langle w_j, x \rangle + w_{0j})}$$

Обучение модели с помощью MLE:

$$\sum_{i=1}^N \log P(y = y_i|x_i, w) \rightarrow \max_{w_1, \dots, w_K}$$

# Кодирование категориальных признаков

Категориальный признак

Binary encoding

Color	Label	colorWhite	colorBlue	colorRed	colorGreen	Binary	Target
White	0	1	0	0	0	8	0.25
Blue	1	0	1	0	0	4	0.25
Red	2	0	0	1	0	2	0.25
Green	3	0	0	0	1	1	0.25

Label encoding

One-hot encoding

Target encoding

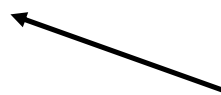
# Наивный Байесовский классификатор

*Формула Байеса:*

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

$$P(y_k|X) = \frac{P(y_k)P(X|y_k)}{P(X)}$$

*Не зависит от класса,  
можно не считать*



Для вектора из  $n$  признаков:

$$P(y_k | X_1, \dots, X_n) = \frac{P(y_k) \prod_{i=1}^n P(X_i | y_k)}{P(X_1, \dots, X_n)}$$

Классификация по правилу:

$$y_k \propto \underset{y_k}{\operatorname{argmax}} P(y_k) \prod_{i=1}^n P(X_i | y_k)$$