Тверской
государственный
технический
университет

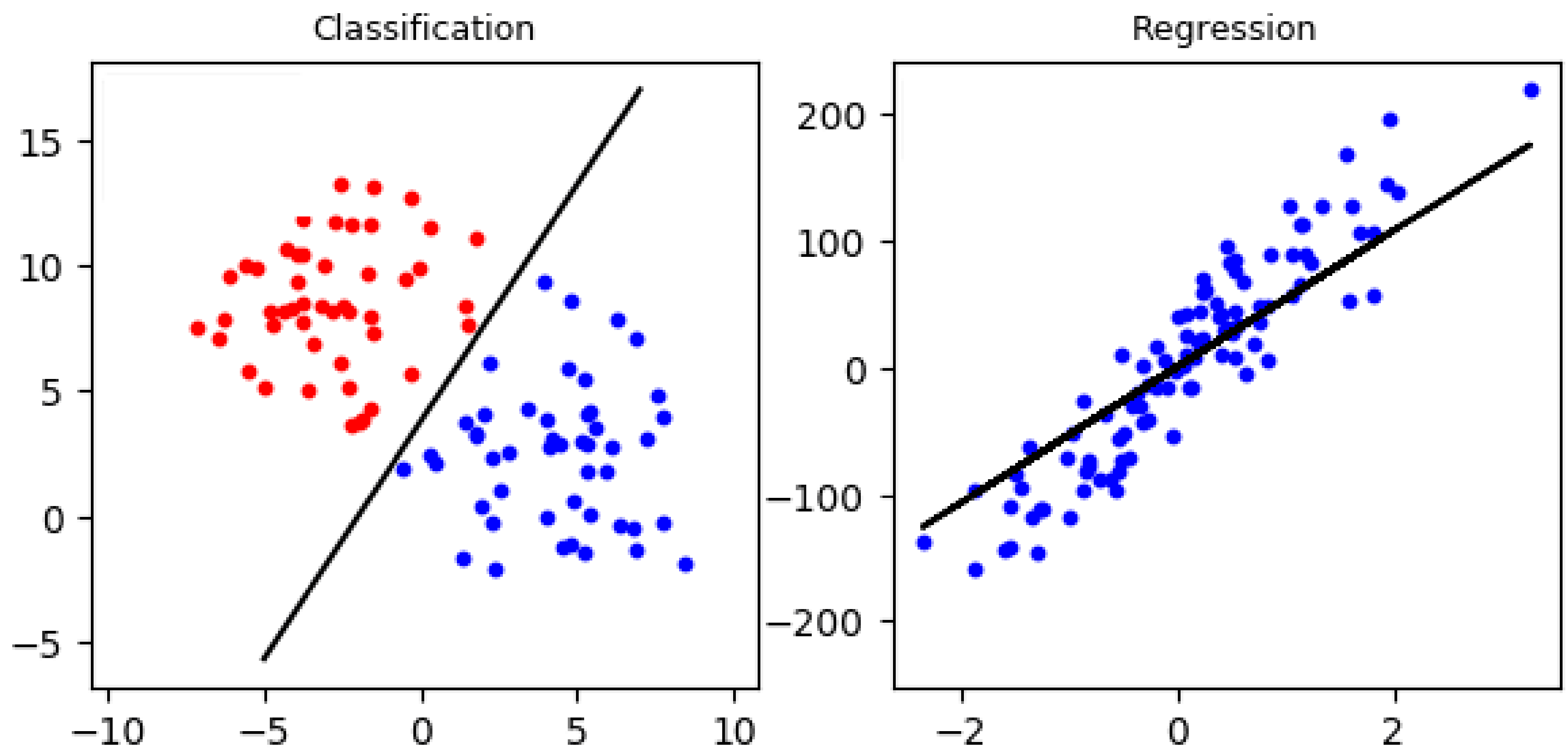
Интеллектуальные информационные системы

Линейные модели классификации
Байесовский классификатор

Шпигарь Андрей Николаевич

Материалы курса доступны по ссылке:
https://github.com/AndreyShpigar/ML-course

2024 г.


Classification
Regression
15
10
5
0
−5
−10
−5
0
5
10
200
100
0
−100
−200
−2
0
2

Линейная модель классификации

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

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

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

<u>Линейная модель классификации:</u>

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

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

Обучение линейного классификатора

- Функционал качества

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

Такой функционал хочется максимизировать

Удобнее решать задачу минимизации, поэтому:

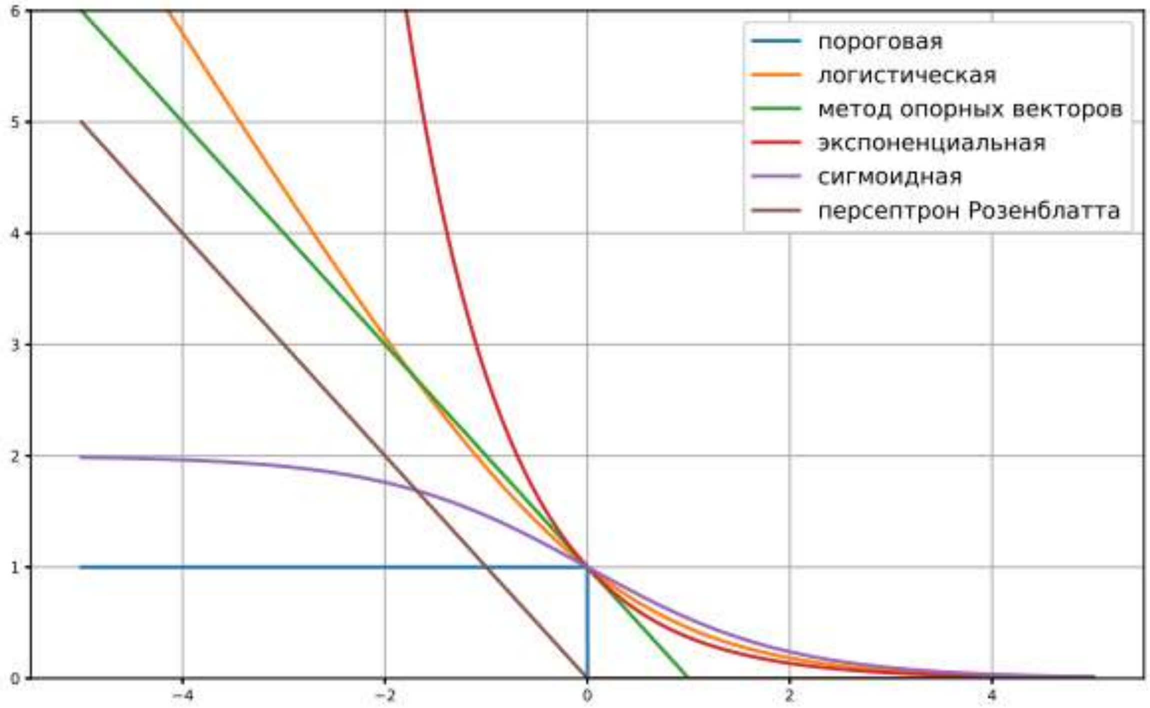
$$\boldsymbol{Q(a,X) = \frac{1}{l}\sum_{i=1}^{l}[a(x_i) \neq y_i] = \frac{1}{l}\sum_{i=1}^{l}[sign\langle w, x_i\rangle \neq y_i] \to \min_{w}}$$

- Концепт «margin» - отступ:

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

$M_i = y_i\langle w, x_i\rangle$ - 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

$$\boldsymbol{log\left(\frac{p}{1-p}\right)}$$ ← *ответ модели*

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

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

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

<u>Правдоподобие</u> – *позволяет понять, насколько вероятно получить данные для значения целевой переменной* ***y*** *при данных* ***X*** *и весах* ***w***.

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

p_i - вероятность, посчитанная из ответов модели

Логарифмическое правдоподобие, хотим чтобы было максимальное

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

Умножим на -1 и будем минимизировать с помощью градиентного спуска

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

Логистическая функция потерь (log-loss)

Многоклассовая классификация

<u>Сведение к набору бинарных задач</u>:

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

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

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

Выбираем самый «уверенный»:

$$\boldsymbol{a(x) = argmax_k(\langle w_k, x \rangle + w_{0k})}$$

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:

$$a(x) = \underset{k \in \{1, \dots, K\}}{argmax} \sum_{i=1}^{K} \sum_{j \neq i} [a_i(x) = k]$$

Многоклассовая логистическая регрессия

Функция Softmax:

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

Вероятность ***k-го*** класса:

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

Обучение модели с помощью MLE:

$$\boldsymbol{\sum_{i=1}^{N} logP(y = y_i|x_i, w) \to \max_{w_1, \dots, w_K}}$$

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:

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

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

Не зависит от класса, можно не считать

Для вектора из ***n*** признаков:

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

Классификация по правилу:

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