

Байесовские Классификаторы

1) Задача классификации

Пусть $X \in \mathbb{R}^d$ — объект, сегодня считаем их случайными векторами.

$y \in \{1, \dots, K\}$ — класс объекта X , который неотъемлемо зависит от X .

Формально: $y = f(X, \varepsilon)$, ε — случайный шум.

Задача классификации

Построить отображение $\hat{y}: \mathbb{R}^d \rightarrow$

Расширенная задача классиф.

Построить отображение $\hat{p}_k(x):$

т.е. \hat{p}_k — оценка вероятности при

$$\hat{p}_1(x) + \dots + \hat{p}_K(x)$$

Тогда $\hat{y}(x) = \arg \max_{k \in \{1, \dots, K\}} \hat{p}_k(x)$

Задача классификации

Построить отображение $\hat{y}: \mathbb{R}^d \rightarrow \{1, \dots, K\}$.

Расширенная задача классиф.

Построить отображение $\hat{p}_k(x): \mathbb{R}^d \rightarrow [0, 1]$,

т.е. \hat{p}_k - оценка вероятности принадлежности x к классу k .

$$\hat{p}_1(x) + \dots + \hat{p}_K(x) = 1.$$

Тогда $\hat{y}(x) = \arg \max_{k \in \{1, \dots, K\}} \hat{p}_k(x)$.

② Байес

Пусть (X_1, \dots, X_n)
где X_i

Применим π -

② Байесовские классиф.

Пусть $(X_1, y_1), \dots, (X_n, y_n)$ — обучающая выборка,
где $X_i \in \mathbb{R}^d$ — вход. объекты
 $y_i \in \{1, \dots, K\}$ — соотв. им классы.

Применим ф-лу Байеса к объекту X и классу $y = k$

$$P(y=k | X=x) = \frac{\pi_k p_k(x)}{\sum_{k=1}^K \pi_k p_k(x)},$$

где $\pi_k = P(y=k)$ — вер-ть класса

$p_k(x) = p(x | y=k)$ — условная плотность X
 $y=k$, т.е. p_k

② Байесовские классиф.

Пусть $(X_1, y_1), \dots, (X_n, y_n)$ — обучающая выборка,

где $X_i \in \mathbb{R}^d$ — атрибуты объектов

$y_i \in \{1, \dots, K\}$ — номер класса.

Применим q -ую Байеса к объекту X и классу $Y = k$:

$$P(Y=k | X=x) = \frac{\pi_k \hat{p}_k(x)}{\sum_{k=1}^K \pi_k \hat{p}_k(x)},$$

где $\pi_k = P(Y=k)$ — априорная вероятность класса

$\hat{p}_k(x) = p(x | Y=k)$ — условная плотность X при условии $Y=k$, т.е. о классе k .

Пусть $\hat{\pi}_k, \hat{p}_k(x)$ — оценки π_k и $p_k(x)$ соотв.

Тогда:
$$\hat{P}_k(x) = \frac{\hat{\pi}_k \hat{p}_k(x)}{\sum_{s=1}^K \hat{\pi}_s \hat{p}_s(x)}$$

это оценка вероятности принагл. x к классу k .

$$\hat{y}(x) = \operatorname{argmax}_{k \in \{1, \dots, K\}} \hat{\pi}_k \hat{p}_k(x)$$

③ Наивный Байес

Наивное предположение:

все признаки внутри каждого класса независимы.

Рассмотрим частный случай бинарных признаков.

$x_i \in \{0, 1\}$

$$\underline{P(X_i = x_i | y_i = y)} = \prod_{j=1}^d \underbrace{P(x_j = x_j | y_i = y)}_{\text{признак}}$$

Обозначим $P(x_{ij} = 1 | y_i = k) = p_{kj}$

Тогда $\underline{P(X_i = x | y_i = k)} = \prod_{j=1}^d p_{kj}^{x_j} (1 - p_{kj})^{1-x_j}$

$$1 \Rightarrow p_{kj}^1 (1 - p_{kj})^0 = p_{kj}$$

②

Пусть

Применим

Обозначим $P(X_{ij} = 1 | y_i = k) = p_{kj}$

Тогда $P(X_i = x | y_i = k) = \prod_{j=1}^d p_{kj}^{x_j} (1 - p_{kj})^{1-x_j}$

$$x_j = 1 \Rightarrow p_{kj}^1 (1 - p_{kj})^0 = p_{kj}$$

$$x_j = 0 \Rightarrow p_{kj}^0 (1 - p_{kj})^1 = 1 - p_{kj}$$

Оценки: $\hat{J}_k = \frac{n_k}{n}$, где $n_k = \sum_{i=1}^n I\{y_i = k\}$

$$\hat{p}_{kj} = \frac{1}{n_k} \sum_{i=1}^n I\{y_i = k, X_{ij} = 1\}$$

↑
кон-во объектов в классе k

Правильно класс

$$\hat{y}(x) =$$

это мне

Правило классификации:

$$\hat{y}(x) = \underset{k \in \{1, \dots, K\}}{\operatorname{argmax}} \left(\log \hat{\pi}_k + \sum_{j=1}^d \left[x_j \log \hat{p}_{kj} + (1-x_j) \log (1-\hat{p}_{kj}) \right] \right)$$

это линейный классификатор.

где $\pi_k = P(Y=k)$ — вер-то класса

$p_k(x) = p(x|Y=k)$ — условная плотность X при условии $Y=k$, т.е. в классе k .

④ Гауссовские классификаторы

Предполагаем, что $p_k(x)$ — плотность $N(a_k, \Sigma_k)$

т.е.
$$p_k(x) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma_k}} \exp\left(-\frac{1}{2} \cdot (x - a_k)^T \Sigma_k^{-1} (x - a_k)\right)$$

a_k — вектор средних $\in \mathbb{R}^d$

Σ_k — матрица ковариаций $\in \mathbb{R}^{d \times d}$

- симметричная
- неотриц. определ.

$$\xi = (\xi_1, \dots, \xi_d)$$

$$\Sigma \xi = \left(\text{cov}(\xi_i, \xi_j) \right)_{i,j=1}^d$$

Оценки параметров:

$$\hat{\pi}_k = \frac{n_k}{n}$$

$$\hat{a}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i$$

$$\hat{\Sigma}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} (x_i - \hat{a}_k)(x_i - \hat{a}_k)^T$$

μ_k, Σ_k

$\left(-\frac{1}{2}\right)$

$$\cdot (x - a_k)^T \Sigma_k^{-1} (x - a_k)$$

Оценки параметров:

- $\hat{\pi}_k = \frac{n_k}{n}$, где $n_k = \sum_{i=1}^n I\{y_i = k\}$. (3 купе)

- $\hat{a}_k = \frac{1}{n_k} \sum_{i=1}^n X_i I\{y_i = k\}$.

- $\hat{\Sigma}_k = \frac{1}{n_k} \sum_{i=1}^n I\{y_i = k\} \cdot \underbrace{(X_i - \hat{a}_k)}_{d \times 1} \underbrace{(X_i - \hat{a}_k)^T}_{1 \times d}$

Получаем
 $\hat{y}(x)$

Это

Получаем классифик:

$$\hat{y}(x) = \arg \max_{k \in \{1, \dots, K\}} \left(\log \hat{\pi}_k - \frac{1}{2} \log \det \hat{\Sigma}_k - \frac{1}{2} (x - \hat{a}_k)^T \hat{\Sigma}_k^{-1} (x - \hat{a}_k) \right)$$

Это квадратичный дискриминантный анализ (QDA).



Four whiteboard sections containing mathematical content:

- Section 1 (Left):** Derivation of the Fourier transform of a rectangular pulse. It shows the integral
$$F(\omega) = \int_{-T/2}^{T/2} e^{-j\omega t} dt = \frac{e^{-j\omega T/2} - e^{j\omega T/2}}{-j\omega} = \frac{2 \sin(\omega T/2)}{\omega}$$
 and the resulting magnitude spectrum $|F(\omega)| = 2T \left| \text{sinc}(\omega T/2) \right|$.
- Section 2:** Derivation of the Fourier transform of a triangular pulse. It shows the integral
$$F(\omega) = \int_{-T}^0 t e^{-j\omega t} dt + \int_0^T (T-t) e^{-j\omega t} dt$$
 and the resulting magnitude spectrum $|F(\omega)| = T^2 \left| \text{sinc}^2(\omega T/2) \right|$.
- Section 3:** Derivation of the Fourier transform of a trapezoidal pulse. It shows the integral
$$F(\omega) = \int_{-T}^0 t e^{-j\omega t} dt + \int_0^T (T-t) e^{-j\omega t} dt$$
 and the resulting magnitude spectrum $|F(\omega)| = T^2 \left| \text{sinc}^2(\omega T/2) \right|$.
- Section 4 (Right):** Derivation of the Fourier transform of a sinc pulse. It shows the integral
$$F(\omega) = \int_{-\infty}^{\infty} \text{sinc}(t) e^{-j\omega t} dt = \text{rect}(\omega)$$
 and the resulting magnitude spectrum $|F(\omega)| = \text{rect}(\omega)$.



Получаем классифик:

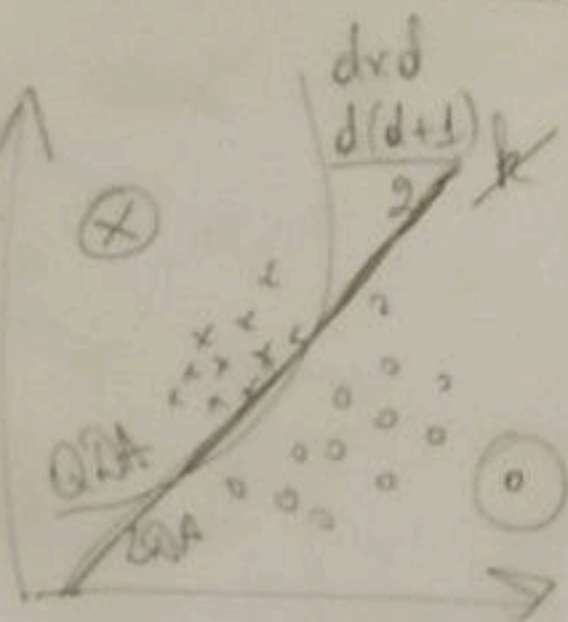
$$\hat{y}(x) = \arg \max_{k \in 1, \dots, K} \left(\log \hat{\pi}_k - \frac{1}{2} \log \det \hat{\Sigma}_k - \frac{1}{2} (x - \hat{\mu}_k)^T \hat{\Sigma}_k^{-1} (x - \hat{\mu}_k) \right)$$

Это квадратичный дискриминантный анализ (QDA).

Линейный дискрим. анализ (LDA)

предполагаем, что $\Sigma_1 = \dots = \Sigma_K = \Sigma$

$$\hat{\Sigma} = \sum_{k=1}^K \hat{\pi}_k \hat{\Sigma}_k$$



Уб в QDA разное
а в LDA

Представим формулу

$$\hat{\pi}_k(x)$$

Рассмотрим:

$$c_k = \frac{1}{2} x^T \hat{\Sigma}_k^{-1} x + \log \hat{\pi}_k, \log \det \dots$$