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 недетерминир. зависит от X.

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

Задача классификации

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

Расширенная задача классиф.

Построить отображение $\hat{p}_k(x)$;

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

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

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

Задача классификации

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

Расширенная задача классиф.

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

т.ч. $\hat{p}_k$ – оценка вероятности принадлежности x к классу k.

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

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

лассу k.

② Байесовские классиф.

Пусть $(X_1, y_1), \ldots, (X_n, y_n)$ – обучающая выборка,

где $X_i \in \mathbb{R}^d$ – случ. объекты

$y_i \in \{1, \ldots, K\}$ – соотв. им класса.

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

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

где $\pi_k = P(y=k)$ – вер-ть класса

$p_k(x) = p(x \mid y=k)$ – условная плотность X

$y = k$, т.е. в

классу k.

② <u>Байесовские классиф.</u>

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

где $X_i \in \mathbb{R}^d$ – случ. объекты

$y_i \in \{1, \dots, K\}$ – соотв. им класса.

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

$$P(y=k \mid X=x) = \frac{\pi_k p_k(x)}{\sum\limits_{k=1}^{K} \pi_k p_k(x)},$$

где $\pi_k = P(y=k)$ – вер-ть класса

$p_k(x) = p(x \mid y=k)$ – условная плотность X при условии $y=k$, т.е. в классе k.

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

Тогда:

$$\hat{P}_k(x) = \frac{\hat{\pi}_k \hat{p}_k(x)}{\sum\limits_{s=1}^{K} \hat{\pi}_s \hat{p}_s(x)}$$

это оценка вероятности принадл. x к классу k.

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

3 Наивный Байес

Наивное предположение:

все признаки внутри каждого класса независимы.

Рассмотрим частный случай бинарных признаков.

$X_{ij} \in \{0, 1\}$

$$P(X_i = x \mid y_i = y) = \prod_{j=1}^{d} P(X_{ij} = x_j \mid y_i = y)$$

i — объект, j — признак

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

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

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

2

Пусть

Примен

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

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

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

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

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

кол-во объектов в классе k

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

Правило клас…

$\hat{y}(x) =$

это лине…

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-ть 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 $\mathcal{N}(a_k, \Sigma_k)$

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

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

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

- симметричная
- неотрицат. опред.

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

$D\xi = \big(\operatorname{cov}(\xi_i, \xi_j)\big)_{i,j}$

Оценки параметров:

- $\hat{\pi}_k = \dfrac{n_k}{n}$,
- $\hat{a}_k = \dfrac{1}{n_k} \sum_{i=1}^{n}$
- $\hat{\Sigma}_k = \dfrac{1}{n_k}$

$$\ldots k, \Sigma_k)$$

$$\ldots\left(-\frac{1}{2}\cdot(x-a_k)^T \Sigma_k^{-1} (x-a_k)\right)$$

Получаем

$$\hat{y}(x \ldots$$

Это …

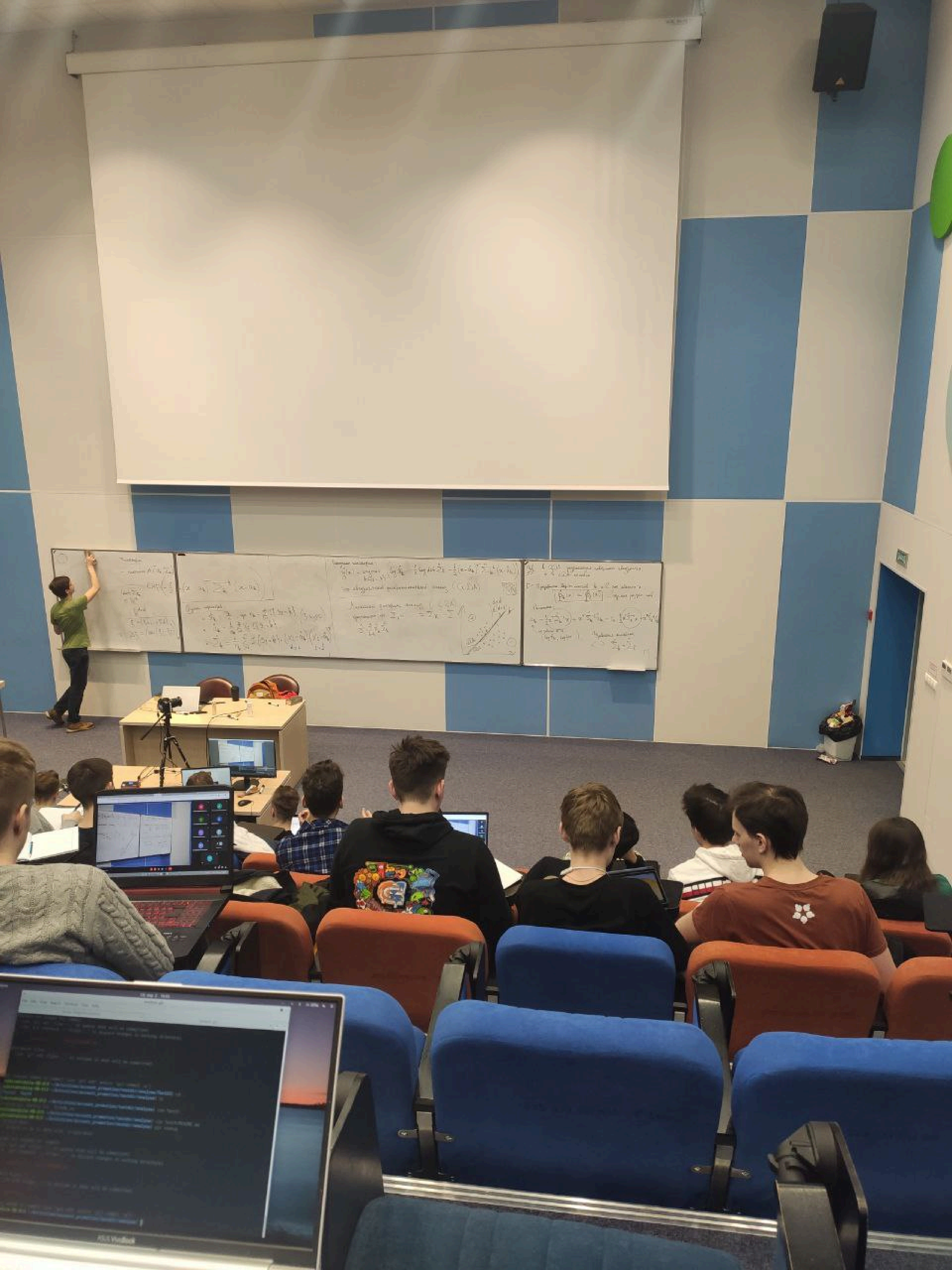
Оценки параметров:

- $\hat{\pi}_k = \frac{n_k}{n}$, где $n_k = \sum\limits_{i=1}^{n} I\{y_i = k\}$. (3 курс)
- $\hat{a}_k = \frac{1}{n_k}\sum\limits_{i=1}^{n} X_i\, I\{y_i = k\}$.
- $\hat{\Sigma}_k = \frac{1}{n_k}\sum\limits_{i=1}^{n} I\{y_i = k\}\cdot \underbrace{(X_i - \hat{a}_k)}_{d\times 1}\underbrace{(X_i - \hat{a}_k)^T}_{1\times d}$

Получаем классифик.:

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

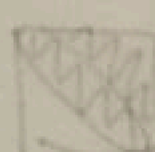
Это квадратичный дискриминантный анализ (QDA).



Получаем классифик.:

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

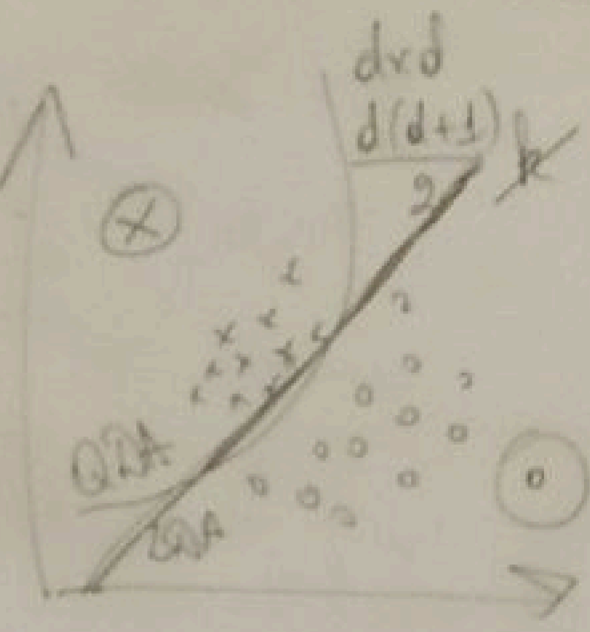
Это квадратичный дискриминантный анализ (QDA)



Линейный дискрим. анализ (LDA)

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

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



Ут. в QDA раздел...
а в LDA ...

▷ Приравняем вер-ти к...

$\boxed{\hat{p}_k(x) \dots}$

Распишем:

$$c_k - \frac{1}{2} x^T \hat{\Sigma}_k^{-1} x + \dots$$

не зависит от x
($\log \hat{\pi}_k$, $\log\det$ ...