



Машинное обучение в науках о Земле

Михаил Криницкий

К.Т.Н., С.Н.С.

Институт океанологии РАН им. П.П. Ширшова

Лаборатория взаимодействия океана и атмосферы и
мониторинга климатических изменений (ЛВОАМКИ)



Задачи классификации

Михаил Криницкий

К.Т.Н., С.Н.С.

Институт океанологии РАН им. П.П. Ширшова

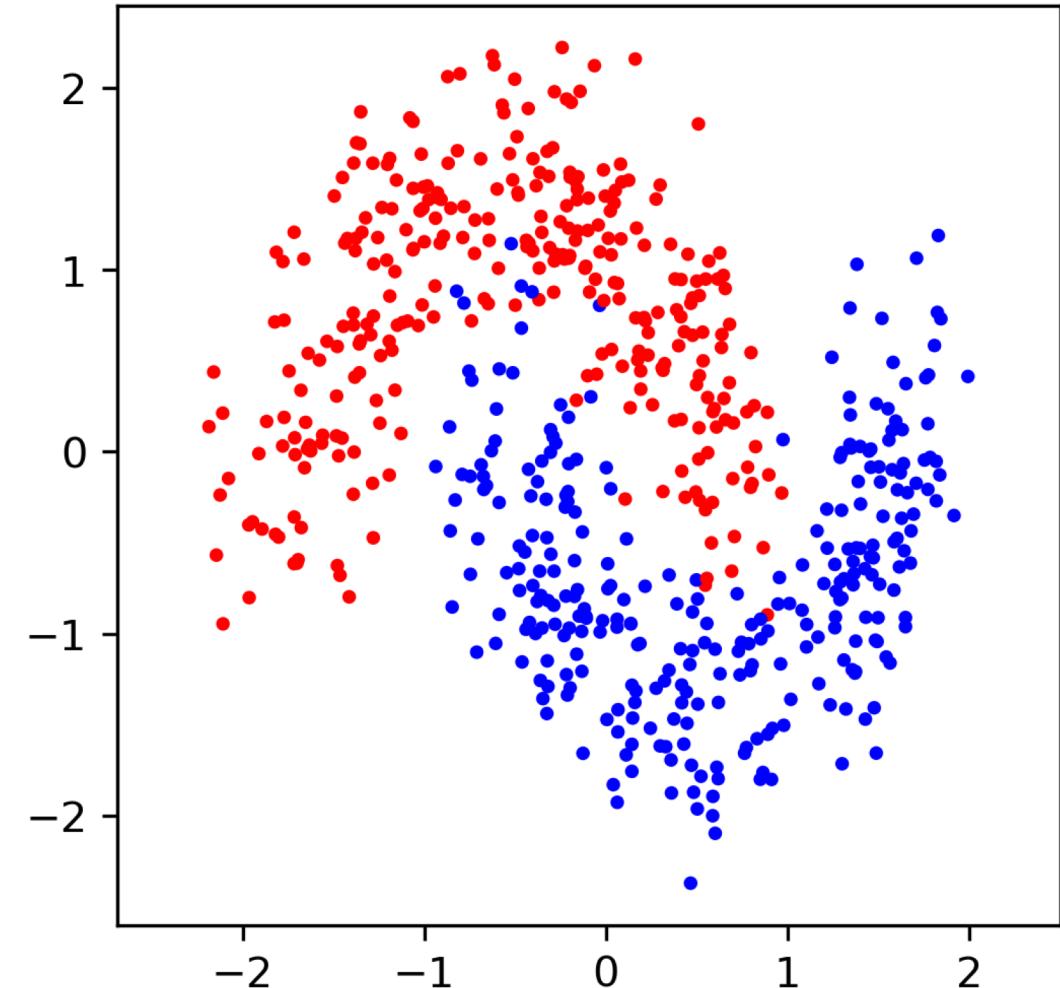
Лаборатория взаимодействия океана и атмосферы и
мониторинга климатических изменений (ЛВОАМКИ)

КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

ЦЕЛЬ: сформулировать задачу (в терминах машинного обучения)

- «Обучение с учителем»
 - восстановление регрессии
 - классификация

ЧТО Я ХОЧУ? – метку класса
«красный или синий?»
(здесь - бинарная классификация)

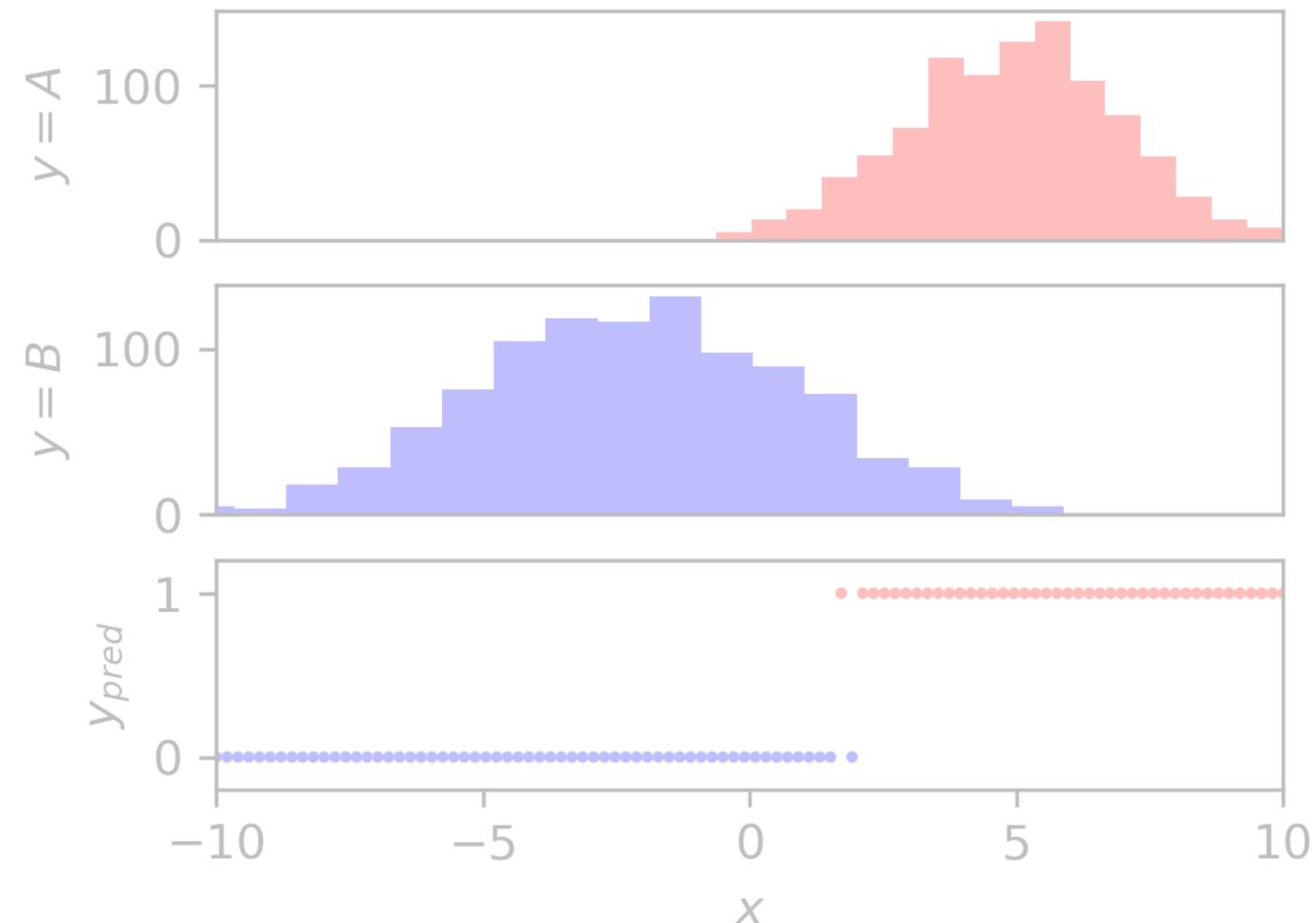


ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Простейший пример: объекты описываются действительным признаком x
 целевая переменная y – бинарная, классы: A , B ; по 1000 экземпляров каждого класса
 пусть для класса $y = A$ значения $x \sim \mathcal{N}(\mu_A, \sigma_A)$, для класса $y = B$ значения $x \sim \mathcal{N}(\mu_B, \sigma_B)$

Подход №1: KNN (метод K ближайших соседей)

1. выбрать K ближайших соседей для нового объекта (! нужно определить меру близости !)
2. осреднить (можно с разными весами) целевую переменную по этим объектам («простое голосование», «majority vote» или «взвешенное голосование», «weighted vote»)
3. считать полученный результат значением целевой переменной на новом объекте



ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

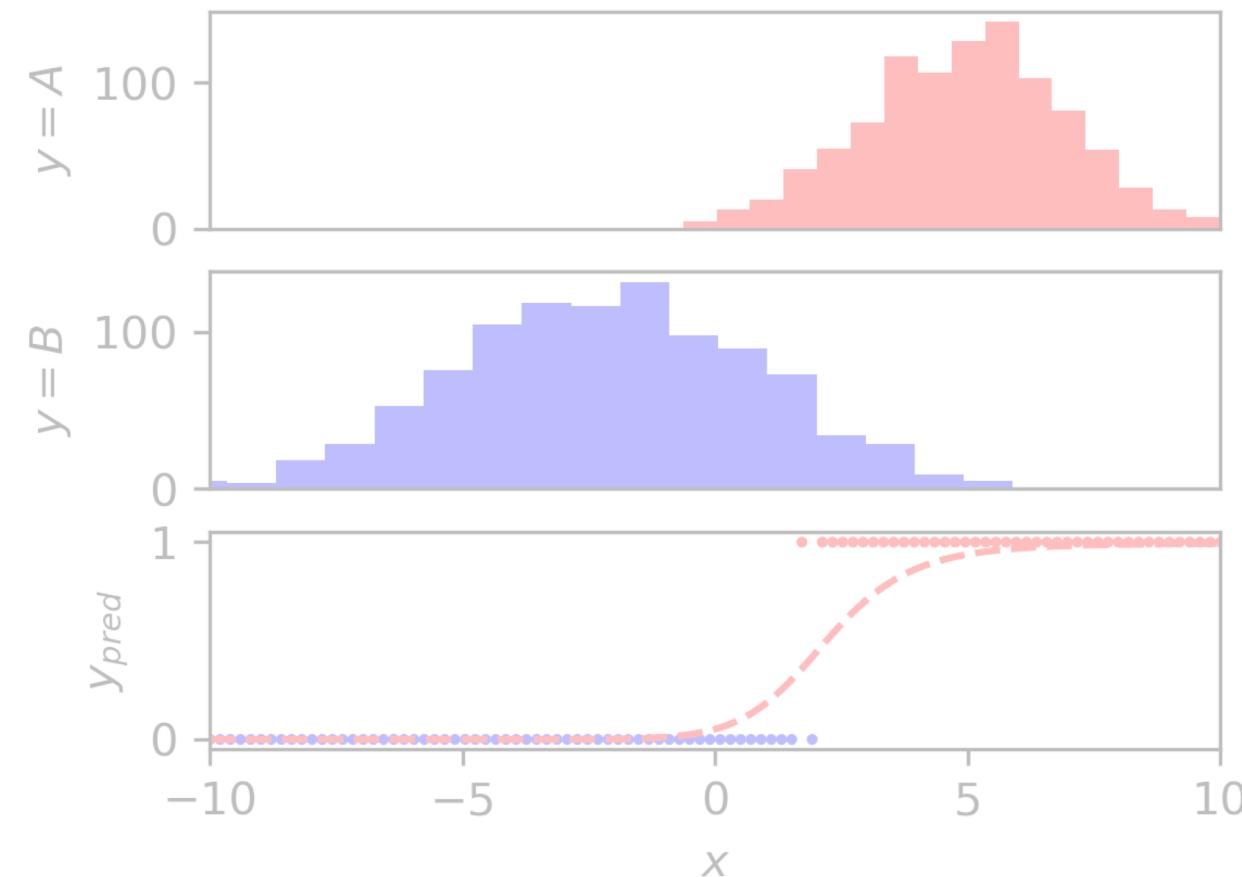
$$P(Y|X) = \frac{P(X|Y) P(Y)}{P(X)}$$

ЕСЛИ нам повезло и мы знаем (или полагаем как допущение в процессе решения) распределения X для каждого из классов $P(X|Y = A)$, $P(X|Y = B)$ etc., - то можно получить **аналитическое решение!**

$$P(Y = B|X = x) = \frac{e^{-\frac{(x+2)^2}{2*9}} * \frac{1}{2}}{e^{-\frac{(x-5)^2}{2*4}} * \frac{1}{2} + e^{-\frac{(x+2)^2}{2*9}} * \frac{1}{2}}$$

«Байесовский классификатор»

(не путать с «naïve bayes»)



ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Naïve bayes classifier, «наивный байесовский классификатор»

$$P(Y|X) \propto P(X|Y) P(Y)$$

можно забыть про $P(X = x)$

$$P(X|Y) = P(x_1, x_2, x_3 \dots x_p | Y) = P(x_2, x_3 \dots x_p | Y, x_1) * P(x_1 | Y) = \dots$$

$$= P(x_1 | Y) * P(x_2 | Y, x_1) * P(x_3 | Y, x_1, x_2) * \dots * P(x_p | Y, x_1, \dots, x_{p-1})$$

наивное предположение: переменные x_i условно независимы:

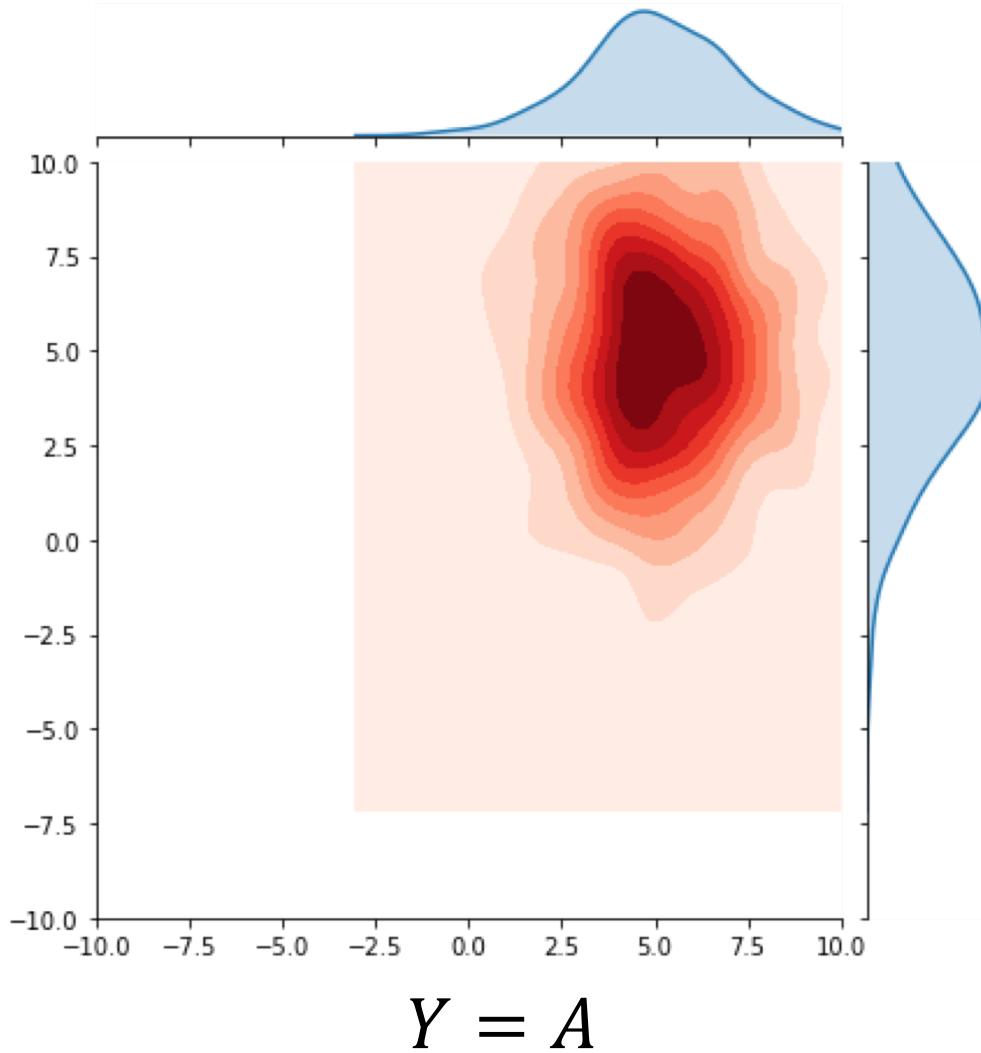
$$P(x_k | Y, x_1, x_2, \dots, x_{k-1}) = P(x_k | Y)$$

Остается оценить распределения $P(x_k | Y)$ для всех k независимо друг от друга – и можно подставлять в оценку вероятности $P(Y|X)$. Оценка распределения может базироваться на предположении об их нормальности, - тогда нужно оценить выборочное среднее (оценка параметра μ) и выборочную дисперсию (оценка параметра σ^2).

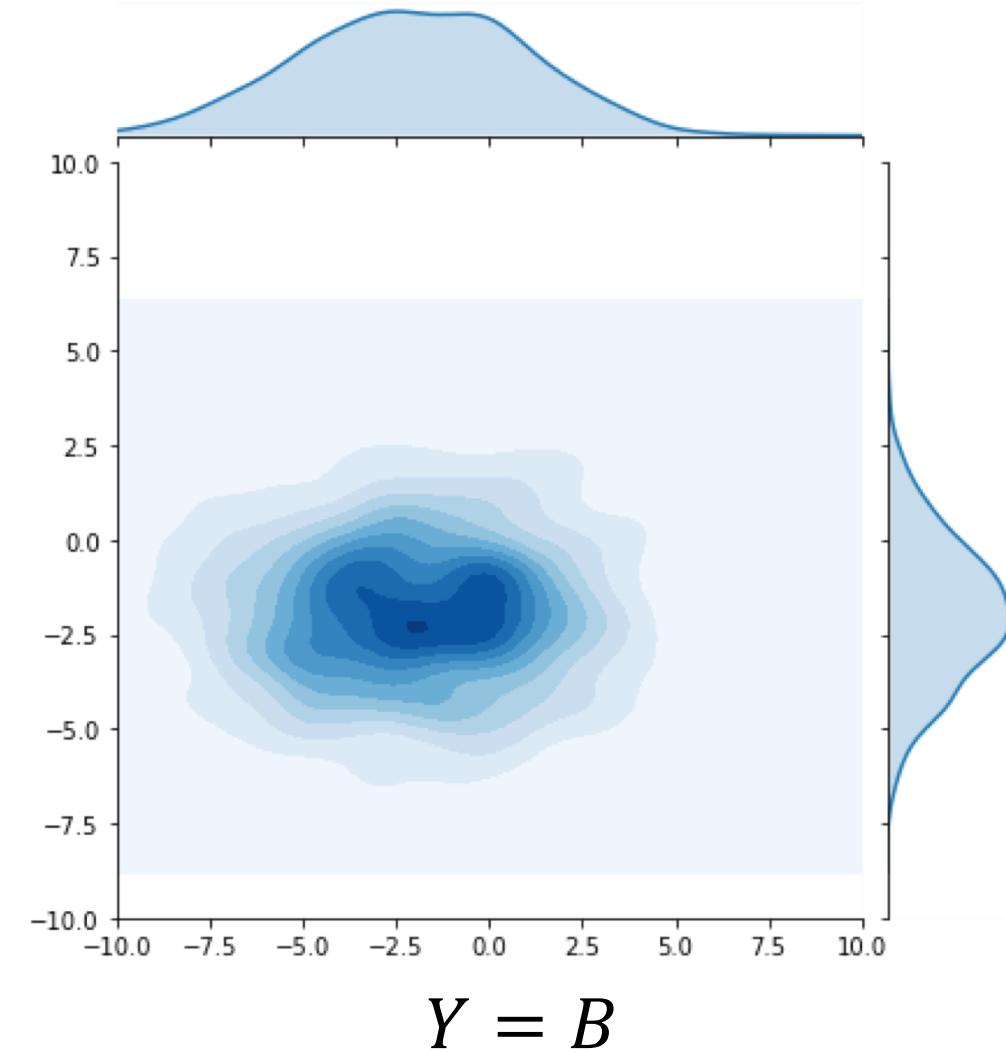
ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Цель: оценить **вероятность** классов ***A*** (класс “1”) и ***B***(класс “0”) для объекта, описываемого значением x .

$$P(Y = k|X = x)$$



$Y = A$

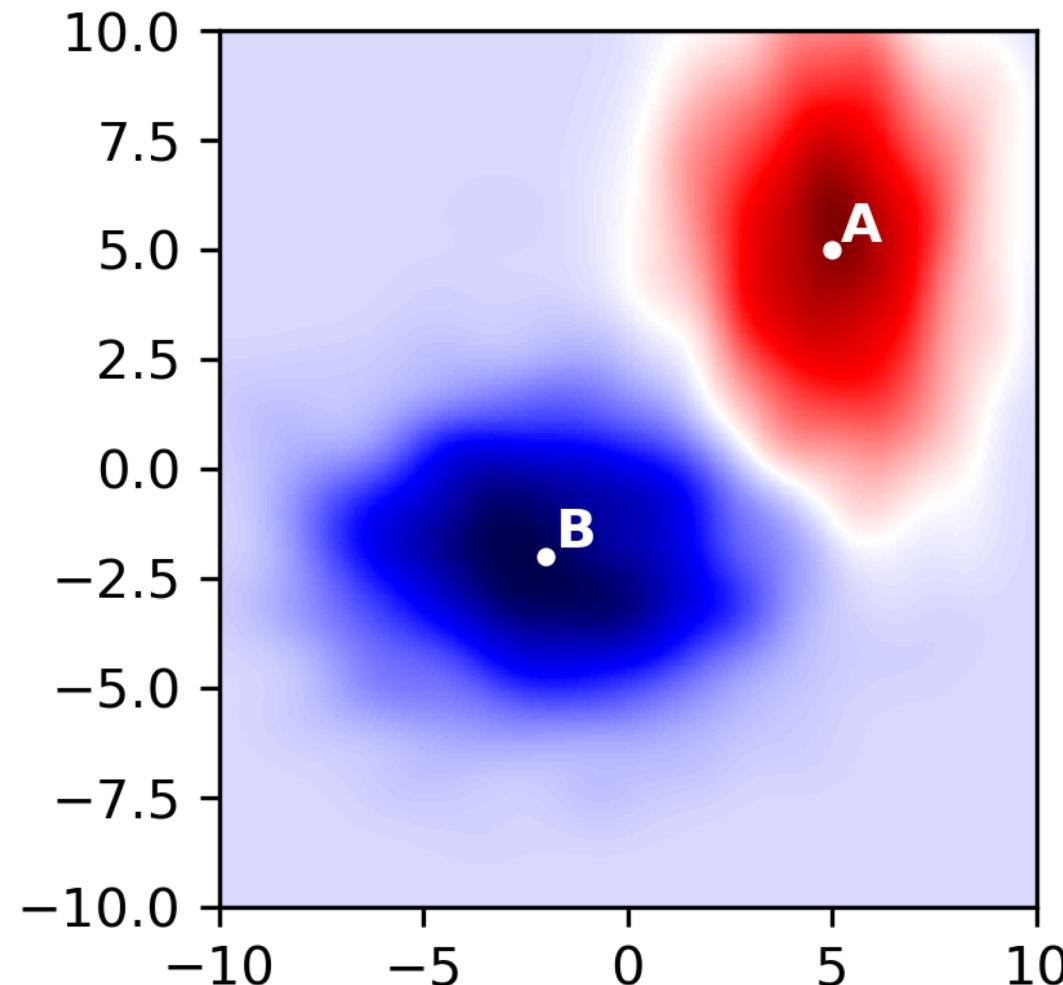


$Y = B$

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Цель: оценить **вероятность** классов **A** (класс “1”) и **B**(класс “0”) для объекта, описываемого значением x .

$$P(Y = k|X = x)$$



ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Naïve bayes classifier, «наивный байесовский классификатор»

$$P(Y|X) \propto P(X|Y) P(Y)$$

можно забыть про $P(X = x)$

$$P(X|Y) = P(x_1, x_2, x_3 \dots x_p | Y) = P(x_2, x_3 \dots x_p | Y, x_1) * P(x_1 | Y) = \dots$$

$$= P(x_1 | Y) * P(x_2 | Y, x_1) * P(x_3 | Y, x_1, x_2) * \dots * P(x_p | Y, x_1, \dots, x_{p-1})$$

наивное предположение: переменные x_i условно независимы:

$$P(x_k | Y, x_1, x_2, \dots, x_{k-1}) = P(x_k | Y)$$

Остается оценить распределения $P(x_k | Y)$ для всех k независимо друг от друга – и можно подставлять в оценку вероятности $P(Y|X)$. Оценка распределения может базироваться на предположении об их нормальности, - тогда нужно оценить выборочное среднее (оценка параметра μ) и выборочную дисперсию (оценка параметра σ^2).

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Naïve bayes classifier, «наивный байесовский классификатор»

$$P(Y|X) \propto P(X|Y) \cdot P(Y)$$

можно забыть про $P(X = x)$

$$\begin{aligned} P(X|Y) &= P(x_1, x_2, x_3 \dots x_p | Y) = P(x_2, x_3 \dots x_p | Y, x_1) * P(x_1 | Y) = \dots \\ &= P(x_1 | Y) * P(x_2 | Y, x_1) * P(x_3 | Y, x_1, x_2) * \dots * P(x_p | Y, x_1, \dots, x_{p-1}) \end{aligned}$$

наивное предположение: переменные x_i условно независимы:

$$P(x_p | Y, x_1, x_2, \dots, x_{p-1}) = P(x_p | Y)$$

Остается оценить распределения $P(x_k | Y)$ для всех k независимо друг от друга – и можно подставлять в оценку вероятности $P(Y|X)$. Оценка распределения может базироваться на предположении об их нормальности, - тогда нужно оценить выборочное среднее (оценка параметра μ) и выборочную дисперсию (оценка параметра σ^2).

NOTE!

LDA (Linear Discriminant Analysis, метод линейного дискриминантного анализа) – то же самое, но в предположении, что (а) распределение $P(x_p | Y)$ - нормальное; (б) для всех классов дисперсия одинакова (матрица ковариаций одна и та же).

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Naïve bayes classifier, «наивный байесовский классификатор»

$$P(Y|X) \propto P(X|Y) P(Y)$$

можно забыть про $P(X = x)$

$$\begin{aligned} P(X|Y) &= P(x_1, x_2, x_3 \dots x_p | Y) = P(x_2, x_3 \dots x_p | Y, x_1) * P(x_1 | Y) = \dots \\ &= P(x_1 | Y) * P(x_2 | Y, x_1) * P(x_3 | Y, x_1, x_2) * \dots * P(x_p | Y, x_1, \dots, x_{p-1}) \end{aligned}$$

наивное предположение: переменные x_i условно независимы:

$$P(x_p | Y, x_1, x_2, \dots, x_{p-1}) = P(x_p | Y)$$

Остается оценить распределения $P(x_k | Y)$ для всех k независимо друг от друга – и можно подставлять в оценку вероятности $P(Y|X)$. Оценка распределения может базироваться на предположении об их нормальности, - тогда нужно оценить выборочное среднее (оценка параметра μ) и выборочную дисперсию (оценка параметра σ^2).

NOTE!

QDA (Quadratic Discriminant Analysis, метод квадратичного дискриминантного анализа) – то же самое, но в предположениях, что (а) распределение $P(x_p | Y)$ - нормальное; (б) для всех компонент x_i дисперсии разные (и здесь σ^2 становится матрицей ковариаций Σ^2). NB с нормальным распределением $P(x_p | Y)$ - это QDA с диагональной Σ^2

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

```
In [ ]: from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
```

```
In [ ]: from sklearn.discriminant_analysis import QuadraticDiscriminantAnalysis
```

```
In [ ]: from sklearn.naive_bayes import GaussianNB
```

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Naïve bayes classifier, «наивный байесовский классификатор»

$$P(Y|X) \propto P(X|Y) \cdot P(Y)$$

Наивное предположение об условной независимости предикторов

Оценка параметров распределений (отдельно для каждого класса, для каждой компоненты признакового описания), подстановка в ф-лу Байеса

$$P(x_p|Y = k, x_1, x_2, \dots, x_{p-1}) = P(x_p|Y = k)$$

$$P(Y = k|X = x) \propto P(Y) * \prod_{j=1}^p P(x_j|Y = k)$$

Вычисление вывода модели

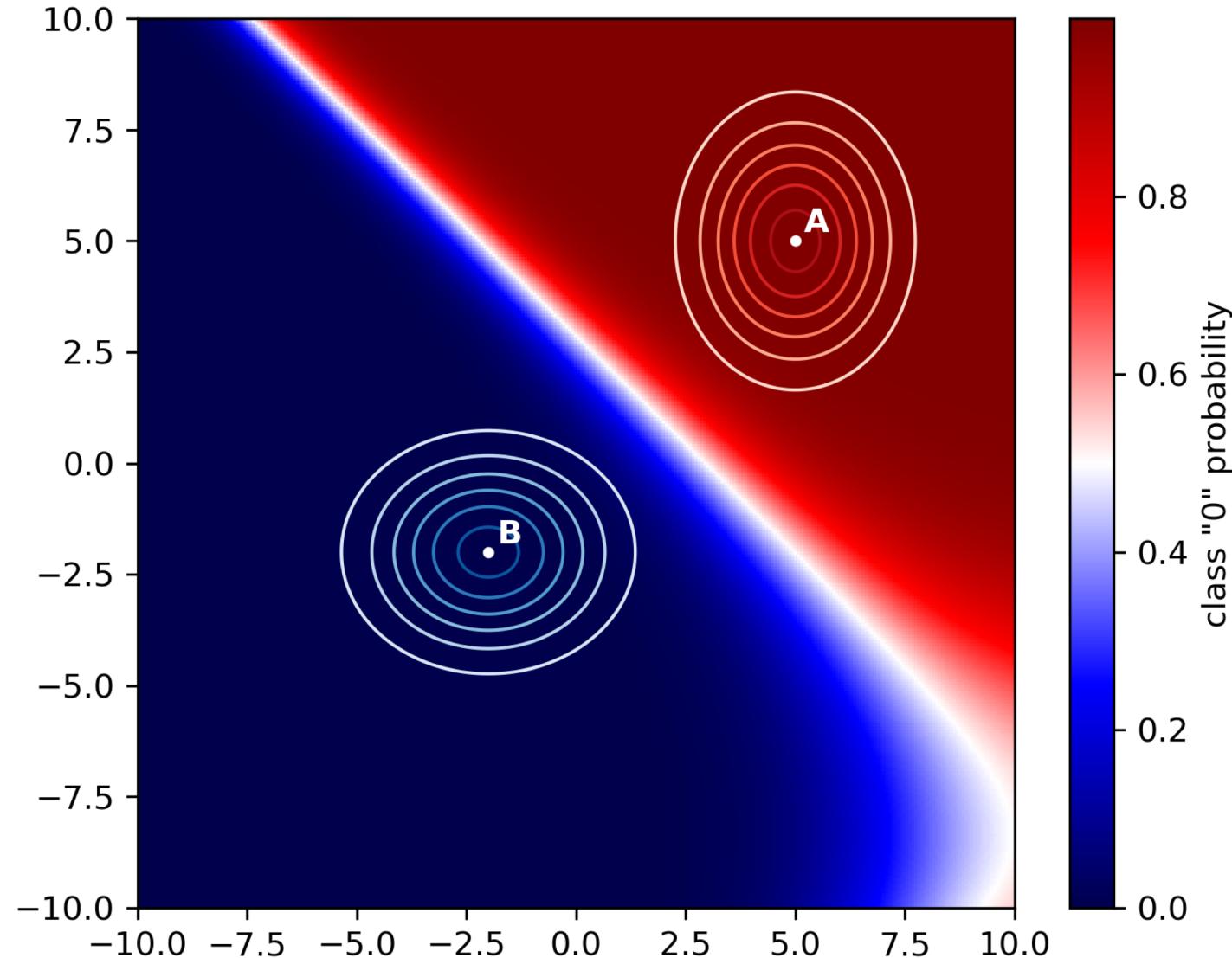
$$\hat{y}(x) = \operatorname{argmax}(P_k(x))$$

Нормировка вероятностей (если нужны оценки вероятностей)

$$P_k(x) = \frac{P(Y = k|X = x)}{\sum_k P(Y = k|X = x)}$$

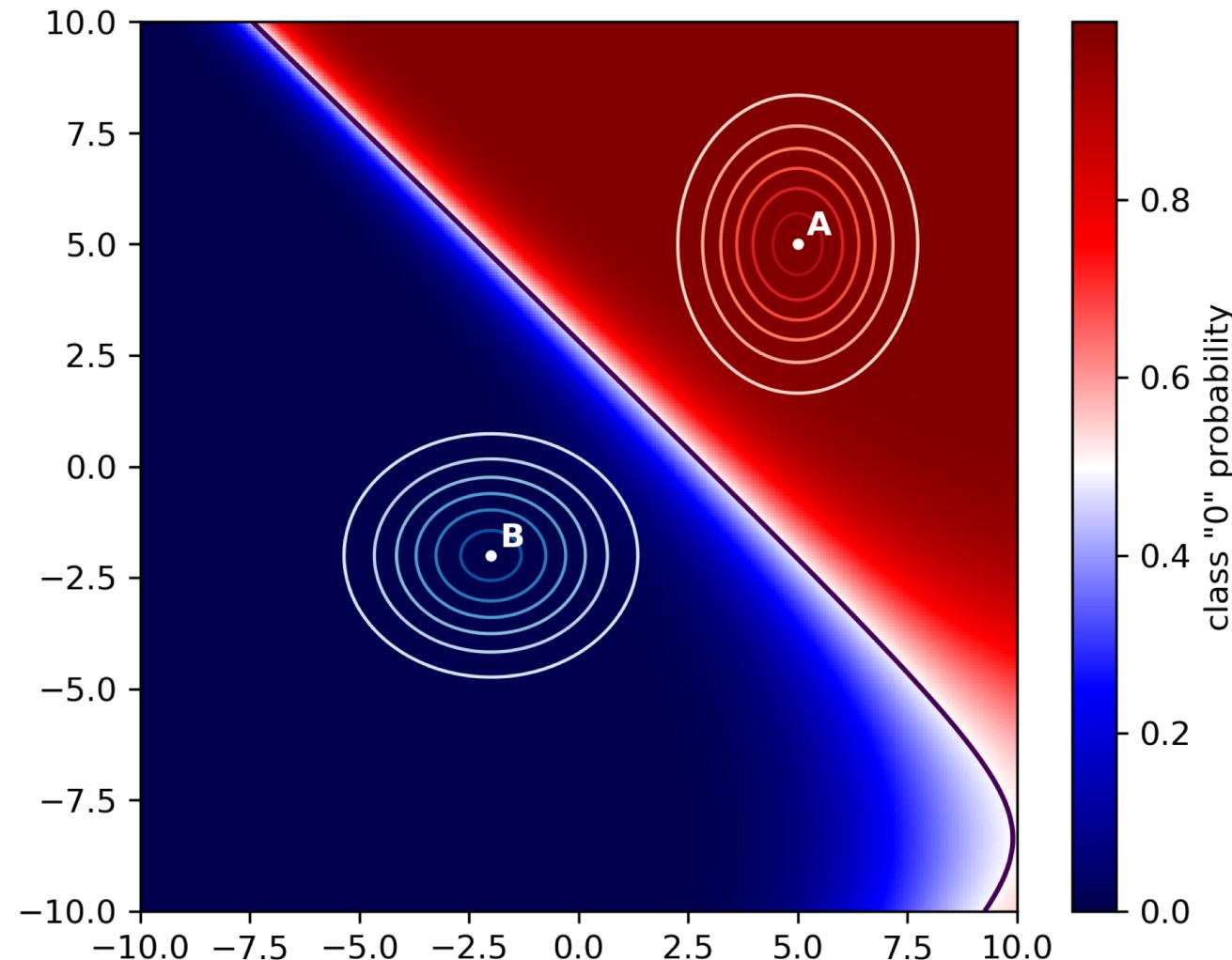
ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Naïve bayes classifier, «наивный байесовский классификатор»

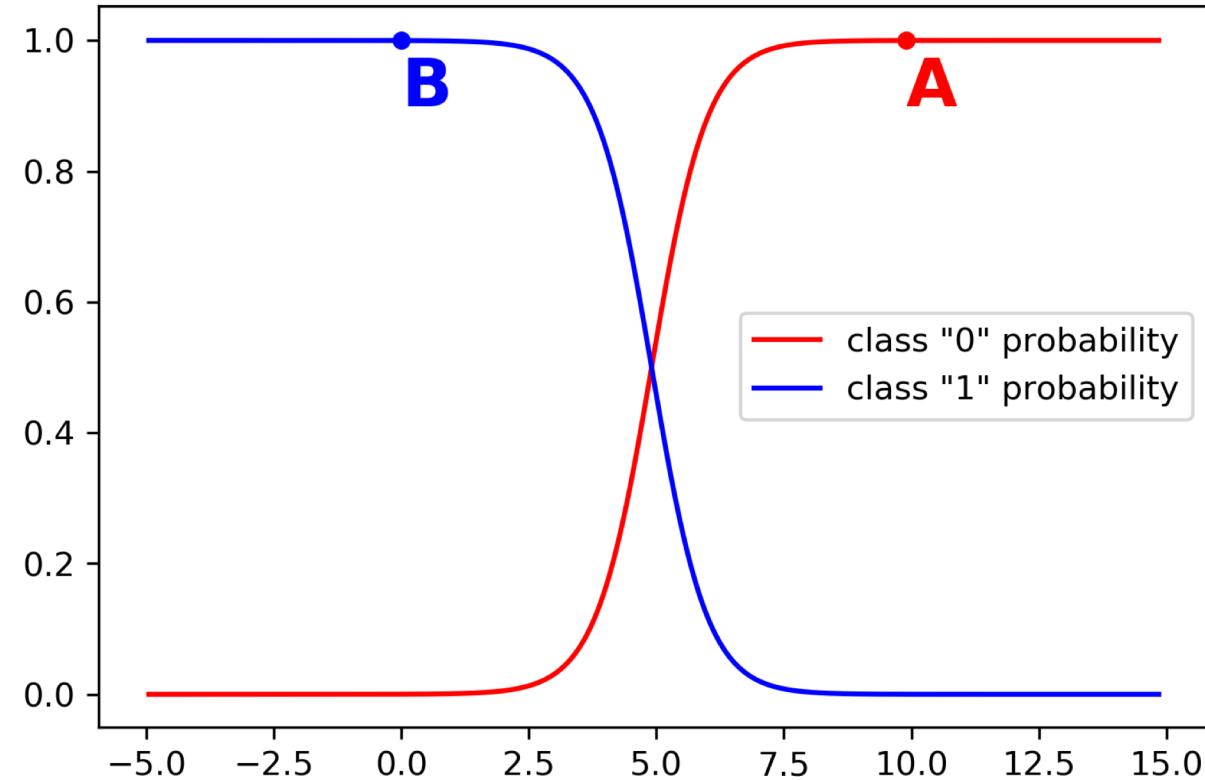
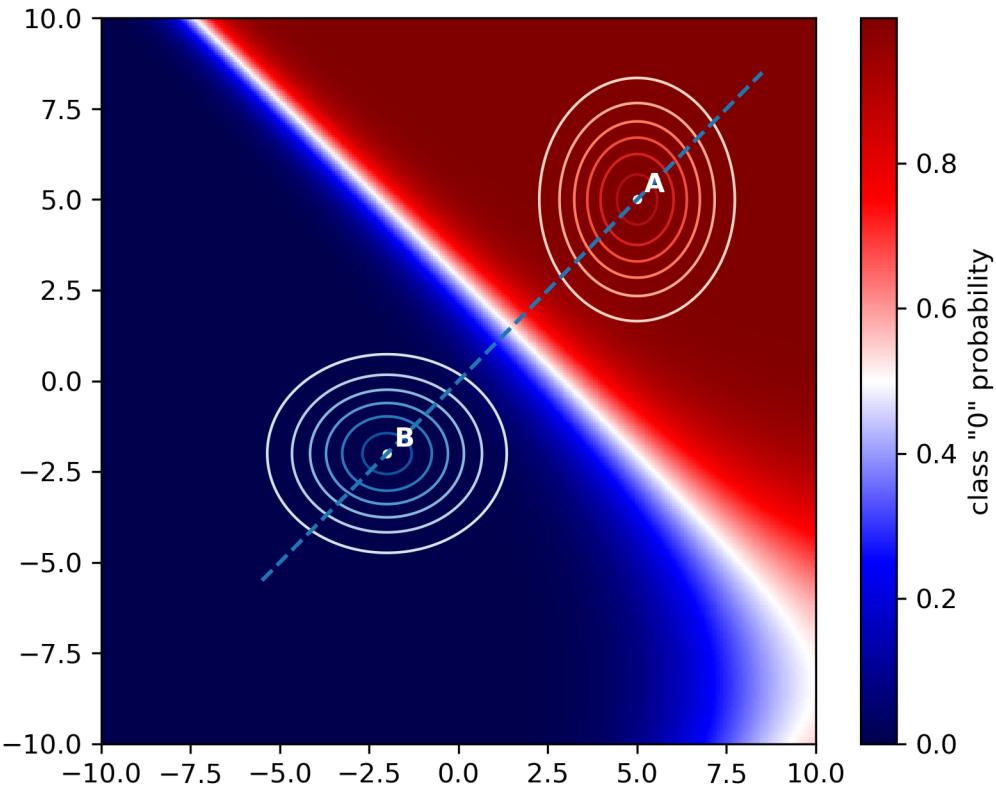


ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Разделяющая поверхность
(на примере результатов наивного байесовского классификатора)



ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации



Может быть, можно аппроксимировать $P(y = 0|x)$ на основании данных обучающей выборки?
Разве не для этого придуман весь подход машинного обучения?

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Логистическая регрессия («logistic regression»)

Линейная регрессия: набор предположений “LINE”

$$y_i \sim \mathcal{N}(\theta \cdot x_i, \sigma^2)$$

$$L(\mathcal{T}) = P(\{y_i, x_i\}) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \theta \cdot x_i)^2}{2\sigma^2}}$$

$$\ell(\mathcal{T}, \theta) = \log L(\mathcal{T}, \theta) = \sum_{i=1}^N \log \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) - \sum_{i=1}^N \frac{(y_i - \theta \cdot x_i)^2}{2\sigma^2}$$

$$\theta^* = \underset{\Theta}{\operatorname{argmax}}(\ell(\mathcal{T}, \theta)) = \underset{\Theta}{\operatorname{argmin}} \sum_{i=1}^N \frac{(y_i - \theta \cdot x_i)^2}{2\sigma^2}$$

Логистическая регрессия:

- (1) Предполагаем, что переменная Y распределена согласно распр.-ю Бернулли с параметром p (зависящий от x_i);
- (2) Предполагаем, что отношение вероятностей классов «1» и «0» соотносятся как линейная функция $f(\theta, x) = \theta \cdot x$

$$Y \sim \mathcal{B}(p(\theta, x)), \quad \log \frac{p(\theta, x)}{1-p(\theta, x)} = \theta \cdot x$$
$$p(\theta, x) = \frac{1}{1 + e^{-\theta \cdot x}}$$

$$L(\mathcal{T}) = \prod_{i=1}^N \left(p(x_i)^{y_i} * (1 - p(x_i))^{(1-y_i)} \right)$$