



Машинное обучение в науках о Земле

Михаил Криницкий

К.Т.Н.,
зав. Лабораторией машинного обучения в науках о Земле МФТИ
с.н.с. Институт океанологии РАН им. П.П. Ширшова



Задачи классификации

Михаил Криницкий

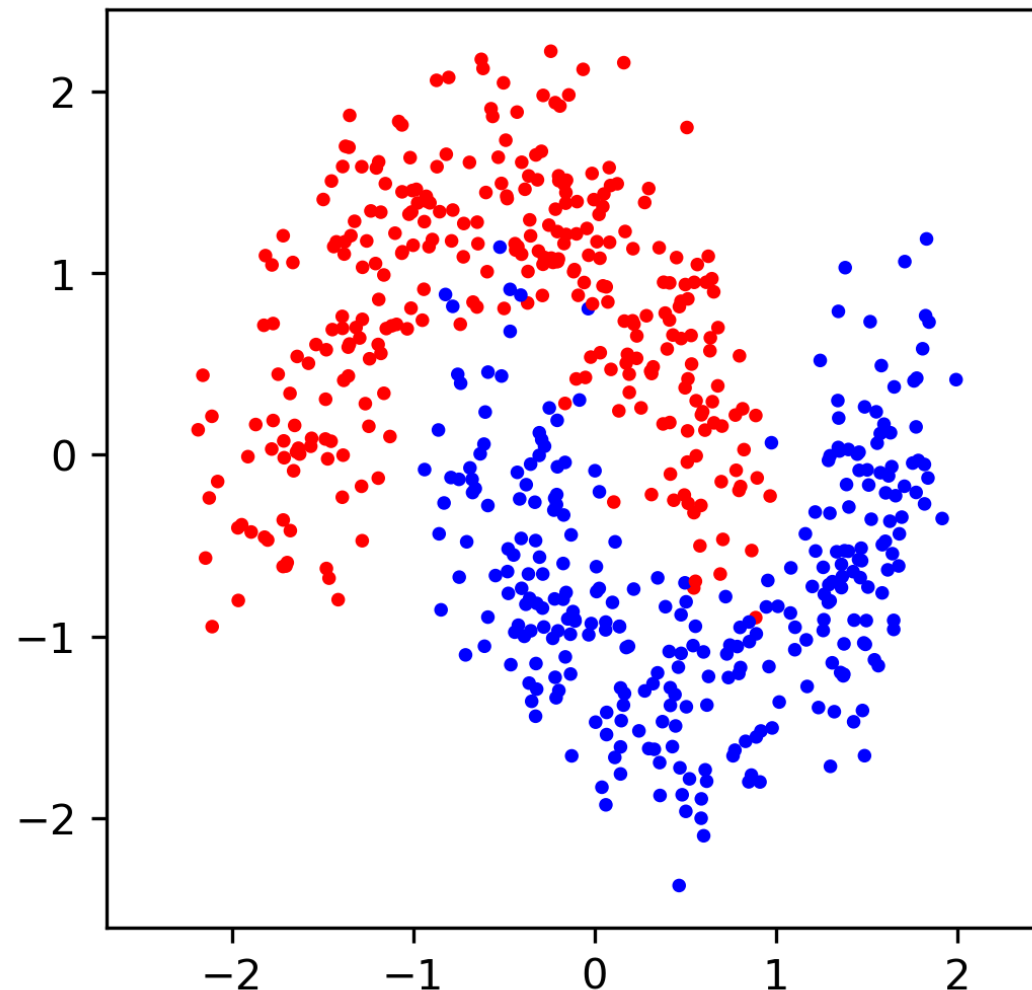
К.Т.Н.,
зав. Лабораторией машинного обучения в науках о Земле МФТИ
с.н.с. Институт океанологии РАН им. П.П. Ширшова

КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

формулируем задачу (в терминах машинного обучения)

- «Обучение с учителем»
 - восстановление регрессии
 - классификация

что я хочу? — метку класса
«красный или синий?»
(бинарная классификация)



ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

цель – метка класса (y – категориальная переменная)

«спам / не-спам»

y – категориальная, бинарная

«мезоциклон / не-мезоциклон»

y – категориальная, бинарная

«кот / собака / лошадь»

y – категориальная, 3 класса

«0 / 1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9»

y – категориальная, 10 классов

«есть дельфин / нет дельфина»

y – категориальная, бинарная

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 $x \sim \mathcal{N}(\mu_B, \sigma_B)$

Базируясь на этих данных, каково должно быть
решение (значение y) при:

$$x = -10$$

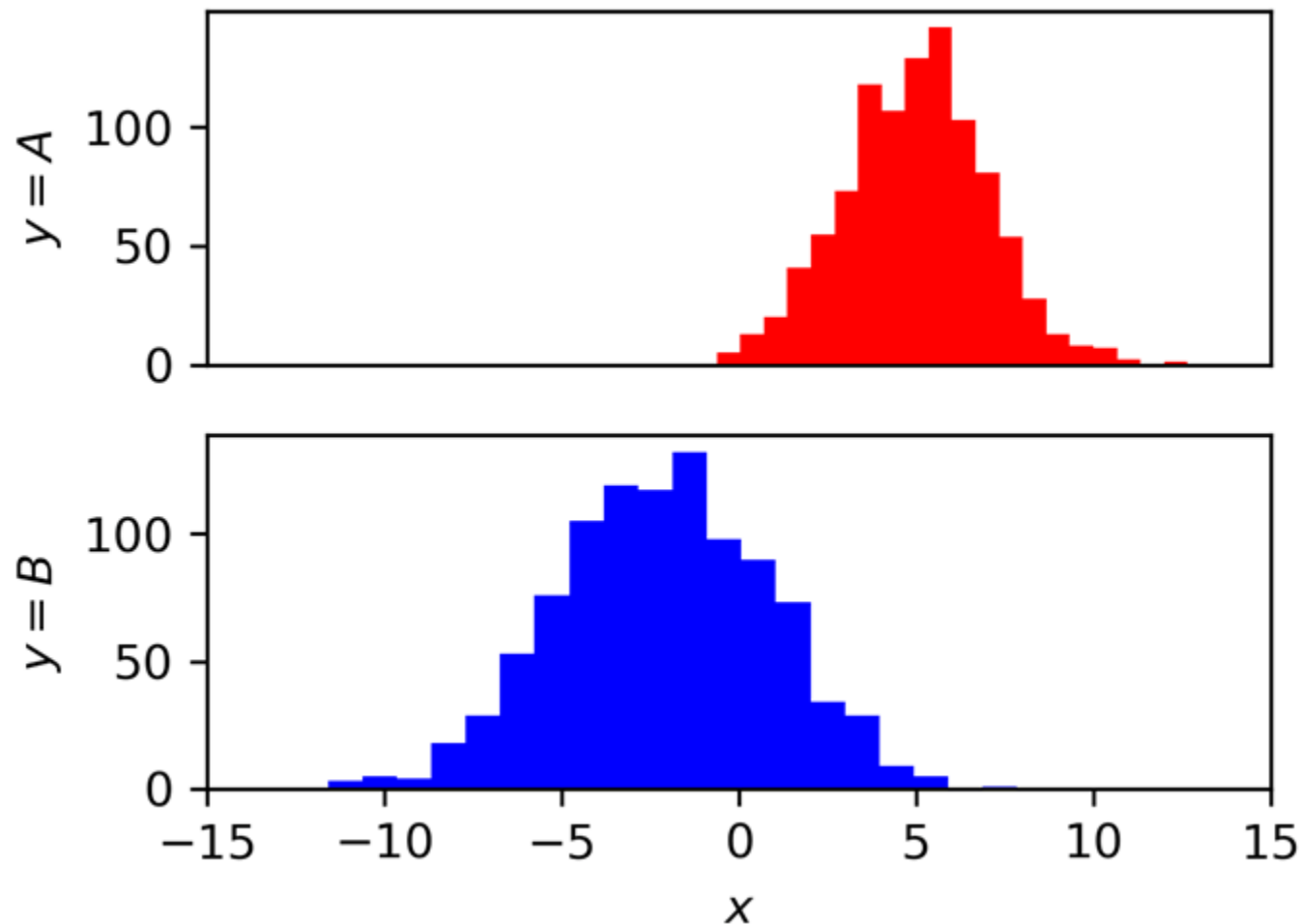
$$x = -5$$

$$x = 2$$

$$x = 5$$

$$x = 10$$

$$x = 15$$

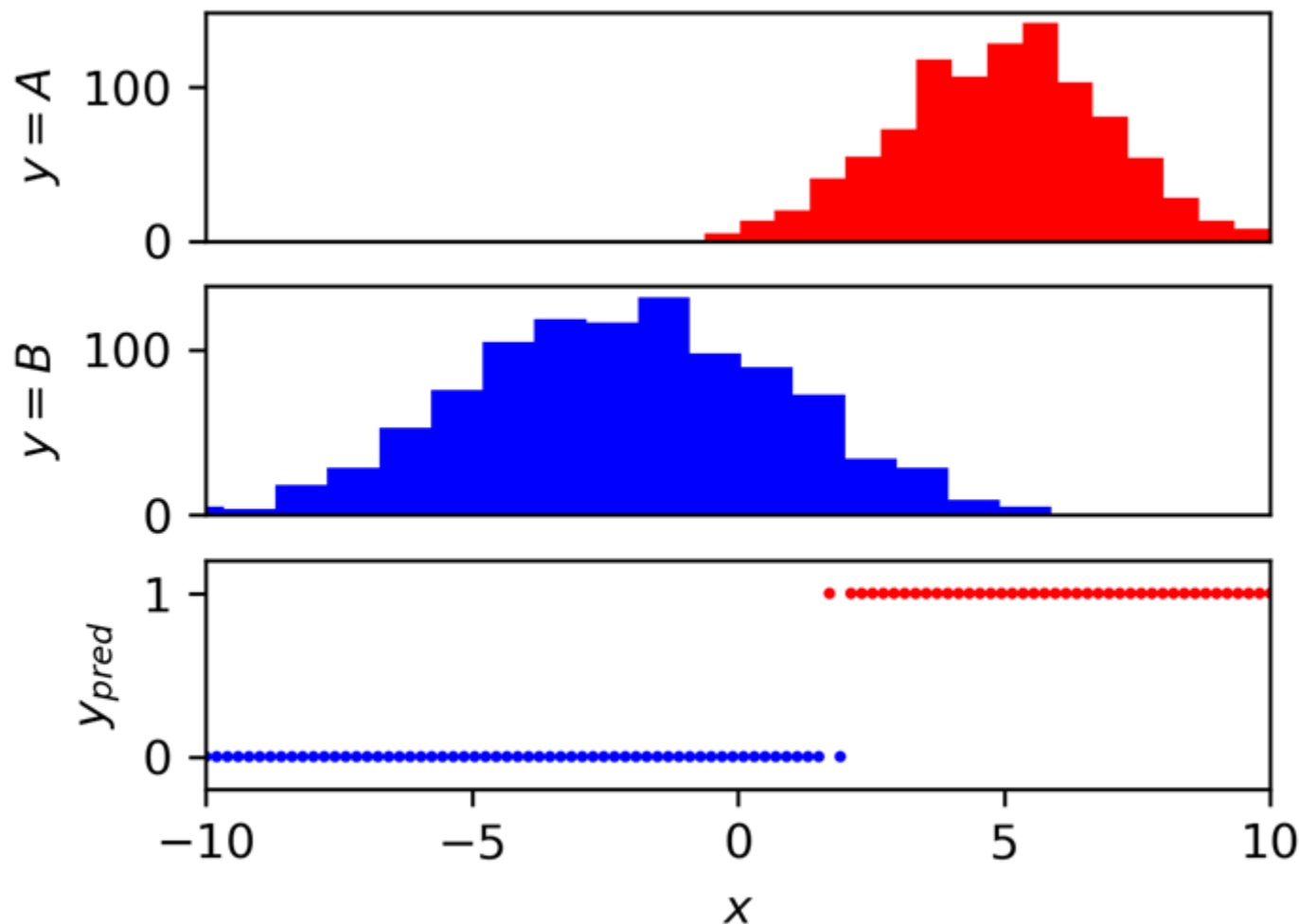


ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

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



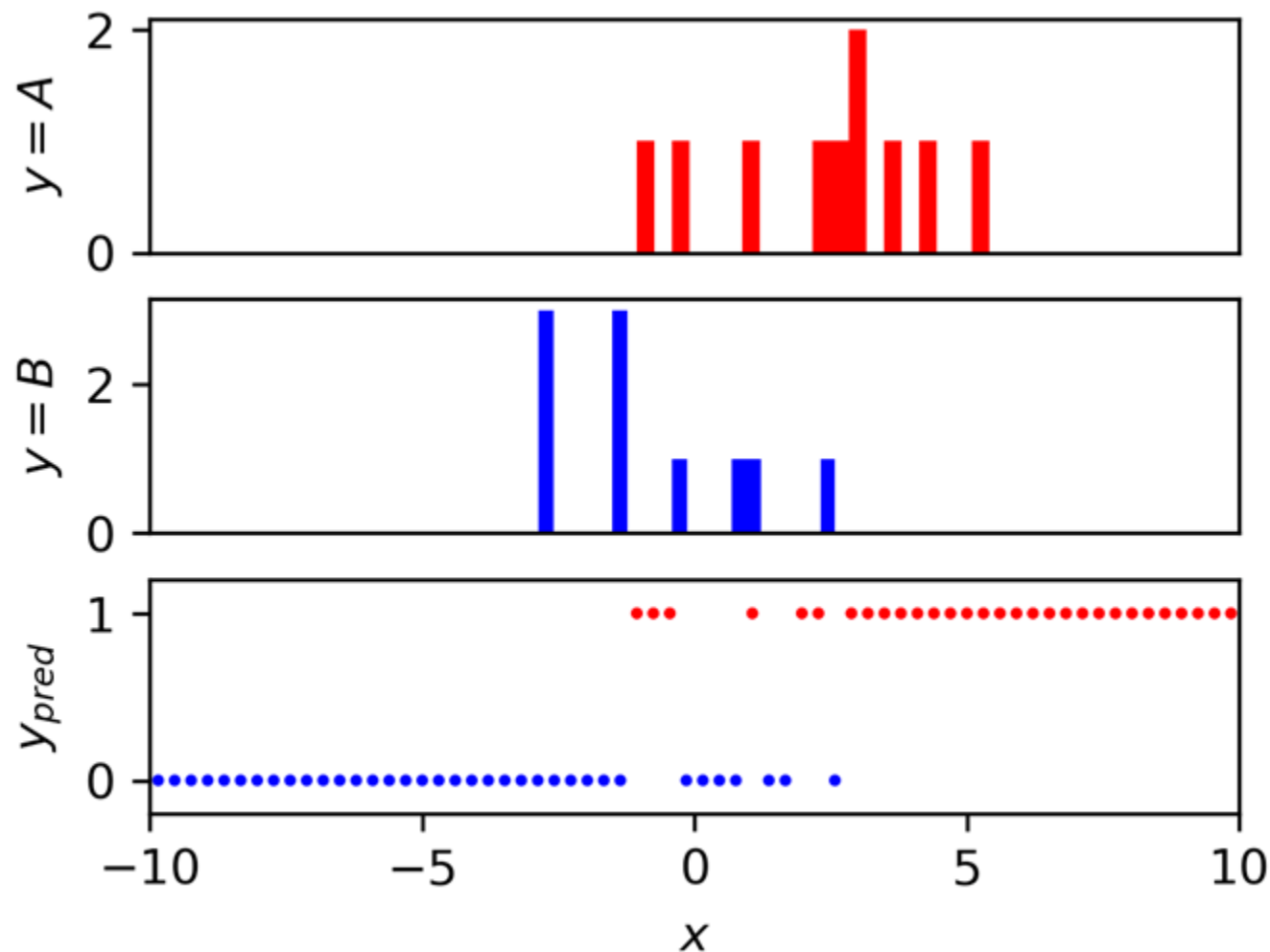
ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Подход №1: **KNN** (метод K ближайших соседей)

- простой
- быстрый
- легко настраивается. Гиперпараметр K регулирует «сложность» модели

А ЧТО ЕСЛИ ДАННЫХ МАЛО?..

- требуется большое количество обучающих данных
- обучающие данные должны быть распределены достаточно плотно в исследуемой области x
- не обобщает закономерности в данных



ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Подход получше – оценить **вероятность** классов **A** и **B** для объекта, описываемого значением x .

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

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Подход получше – оценить **вероятность** классов A и B для объекта, описываемого значением x .

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

Кстати, если нужно принять решение относительно значения Y при определенном значении x_i , помни, что $P(x_i)$ – константа, которую можно не учитывать при сравнении $P(Y = \textcolor{red}{A}|X = x_i)$ и $P(Y = \textcolor{blue}{B}|X = x_i)$

$$P(X) = \sum_{y_i} P(X|Y = y_i)P(Y = y_i)$$

формула полной вероятности

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Подход получше – оценить **вероятность** классов A и B для объекта, описываемого значением x .

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

Кстати, если нужно **принять решение** относительно значения Y при определенном значении x_0 , помни, что $P(x_0)$ – константа, которую можно не учитывать при сравнении $P(Y = A|X = x_0)$ и $P(Y = B|X = x_0)$

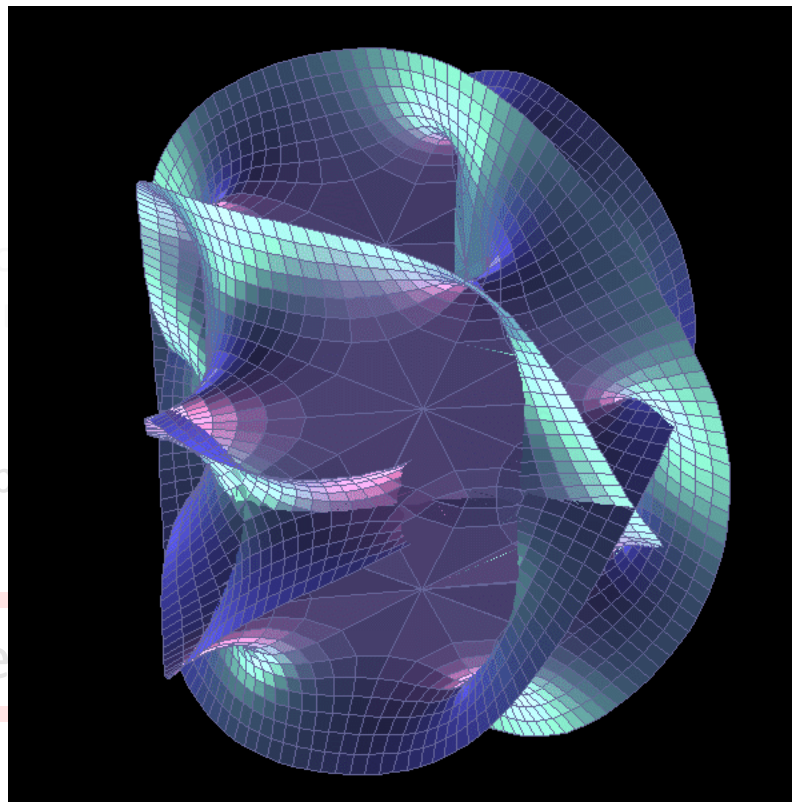
ЕСЛИ нам повезло и МЫ ЗНАЕМ* распределения X для каждого из классов $P(X|Y = A)$, $P(X|Y = B)$ etc., - то можно получить **аналитическое решение!**

И это решение будет ЛУЧШИМ из всех возможных.

*что значит «мы знаем распределения $P(X|Y = A)$, $P(X|Y = B)$ »?

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Подход получше – оценить **вероятность** классов A и B для объекта, описываемого значением x .



Кстати, если нужно принять решение с минимальными потерями, то $P(x_0)$ – константа, которую можно не учитывать.

В данном значении x_0 , помни, что $P(x_0) = P(Y = B | X = x_0)$

ЕСЛИ нам повезло и МЫ ЗНАЕМ* распределения, то можно получить **аналитическое решение!**

$P(X|Y = A)$, $P(X|Y = B)$ etc., - то можно

И это реше

ВОЗМОЖНЫХ.

*что значит «мы знаем распределения $P(X|Y = A)$, $P(X|Y = B)$ »?

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

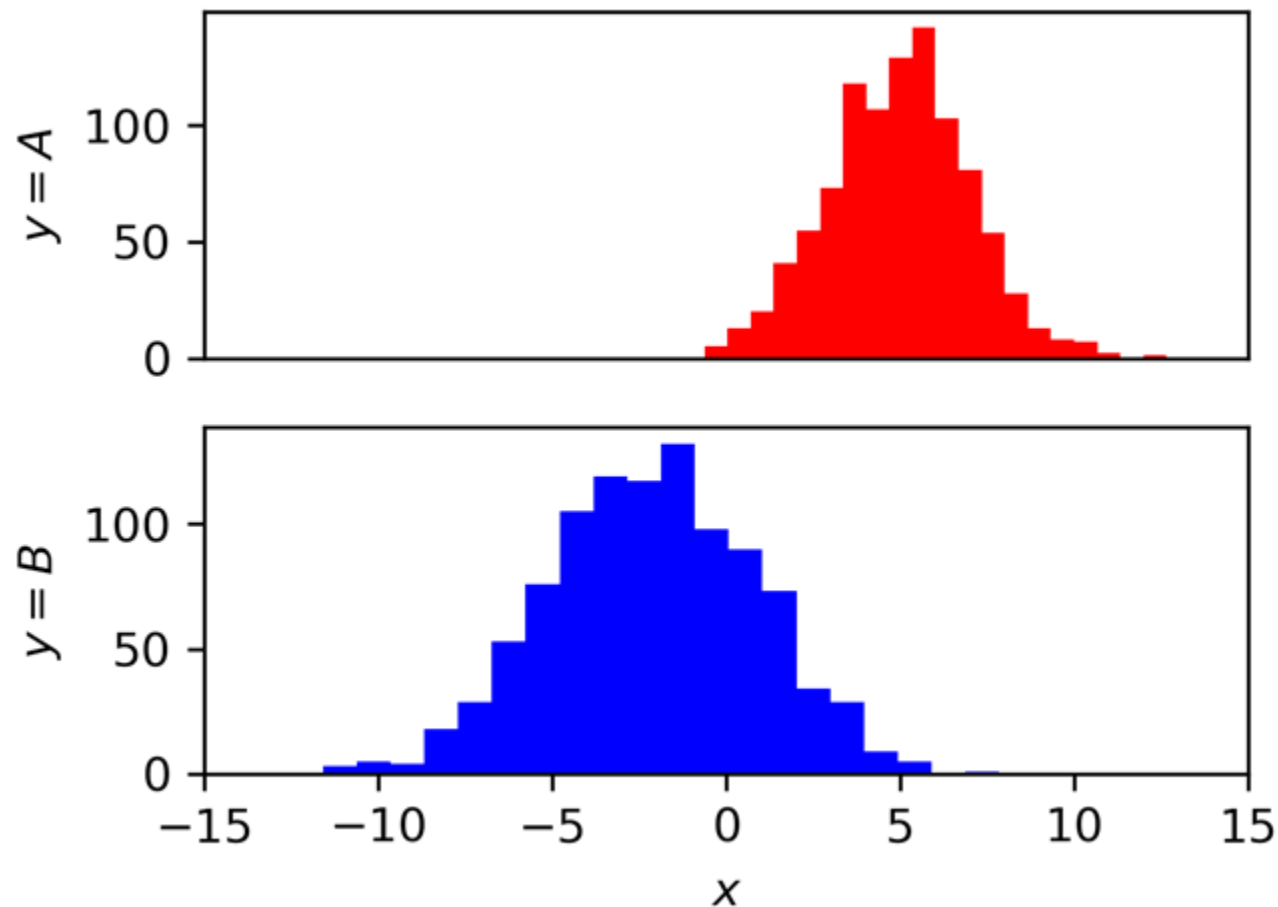
$$P(Y|X) = \frac{P(X|Y) P(Y)}{P(X)}$$

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!**

объекты описываются действительным признаком x
целевая переменная y – бинарная

пусть для класса $y = A$ значения $x \sim \mathcal{N}(\mu_A, \sigma_A)$, для
класса $y = B$ значения $x \sim \mathcal{N}(\mu_B, \sigma_B)$

$$\begin{aligned}\mu_A &= 5 \\ \mu_B &= -2 \\ \sigma_A &= 2 \\ \sigma_B &= 3\end{aligned}$$



ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

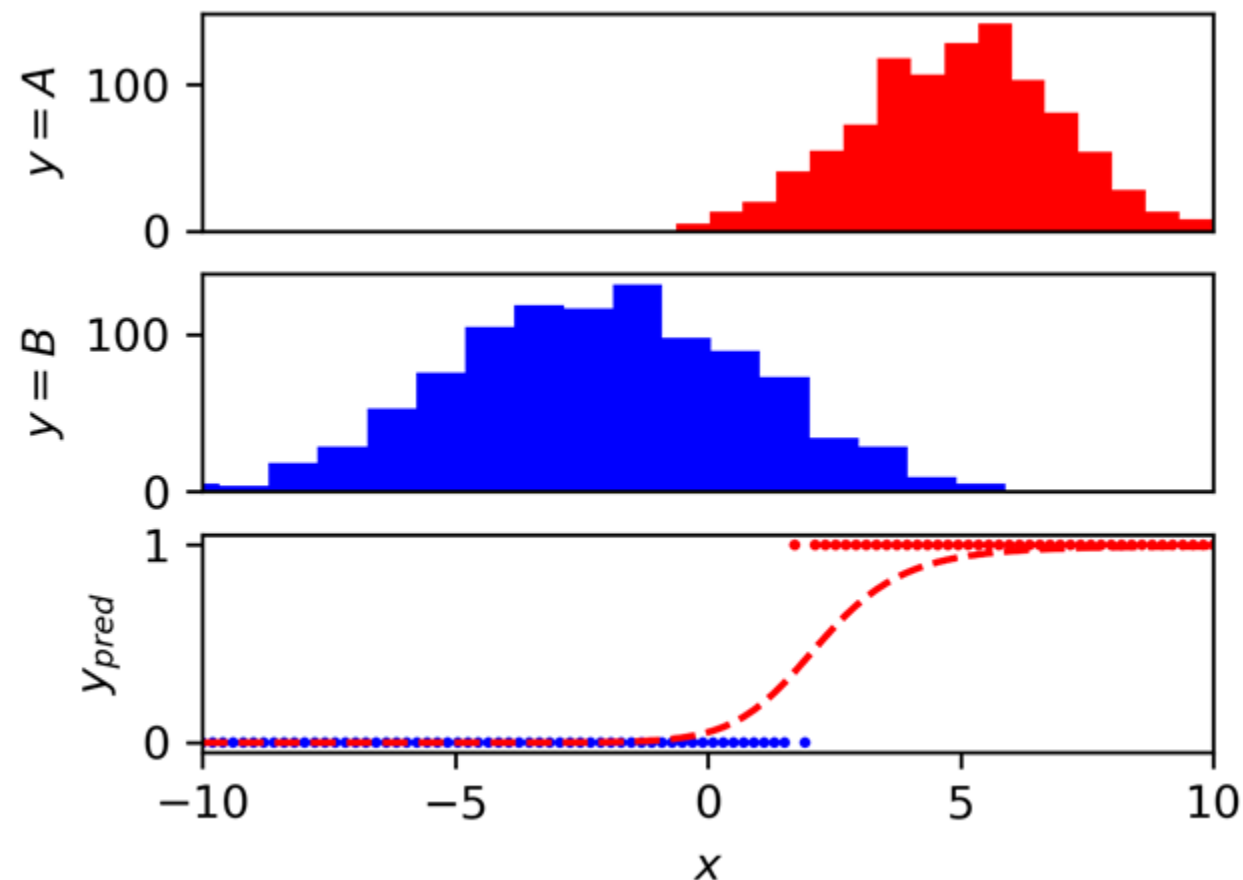
$$P(Y|X) = \frac{P(X|Y) P(Y)}{P(X)}$$

ЕСЛИ нам повезло и МЫ ЗНАЕМ (или полагаем как допущение в процессе решения) распределения X для каждого из классов $P(X|Y = \textcolor{red}{A})$, $P(X|Y = \textcolor{blue}{B})$ etc., - то можно получить **аналитическое решение!**

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

«Байесовский классификатор»

(не путать с «naïve bayes»)



ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Подход получше – оценить **вероятность** классов A и B для объекта, описываемого значением x .

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

ЕСЛИ нам повезло и мы знаем распределения X для каждого из классов $P(X|Y = A)$, $P(X|Y = B)$ etc.,
то можно получить **аналитическое решение!**

И это решение будет **ЛУЧШИМ** из всех возможных.

А ЧТО ЕСЛИ НАМ НЕ ПОВЕЗЛО? *

* чаще всего так и бывает

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Подход получше – оценить **вероятность** классов A и B для объекта, описываемого значением x .

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

ЕСЛИ нам повезло и мы знаем распределения X для каждого из классов $P(X|Y = A)$, $P(X|Y = B)$ etc.,
то можно получить **аналитическое решение!**

И это решение будет **ЛУЧШИМ** из всех возможных.

А ЧТО ЕСЛИ НАМ НЕ ПОВЕЗЛО???

**Надо как-то оценить распределения $P(X|Y = A)$, $P(X|Y = B)$
руководствуясь данными, которые у нас на руках**

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Naïve bayes classifier, «наивный байесовский классификатор»

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

можно забыть про $P(X = x)$

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


ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Naïve bayes classifier, «наивный байесовский классификатор»

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

можно забыть про $P(X = x)$

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

$$P(x_1, \dots) = P(\dots | x_1) * P(x_1)$$


ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Naïve bayes classifier, «наивный байесовский классификатор»

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

можно забыть про $P(X = x)$

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

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

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Naïve bayes classifier, «наивный байесовский классификатор»

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

можно забыть про $P(X = x)$

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

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

наивное предположение: переменные x_i условно независимы:

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

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

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 $P(Y|X)$.

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:

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

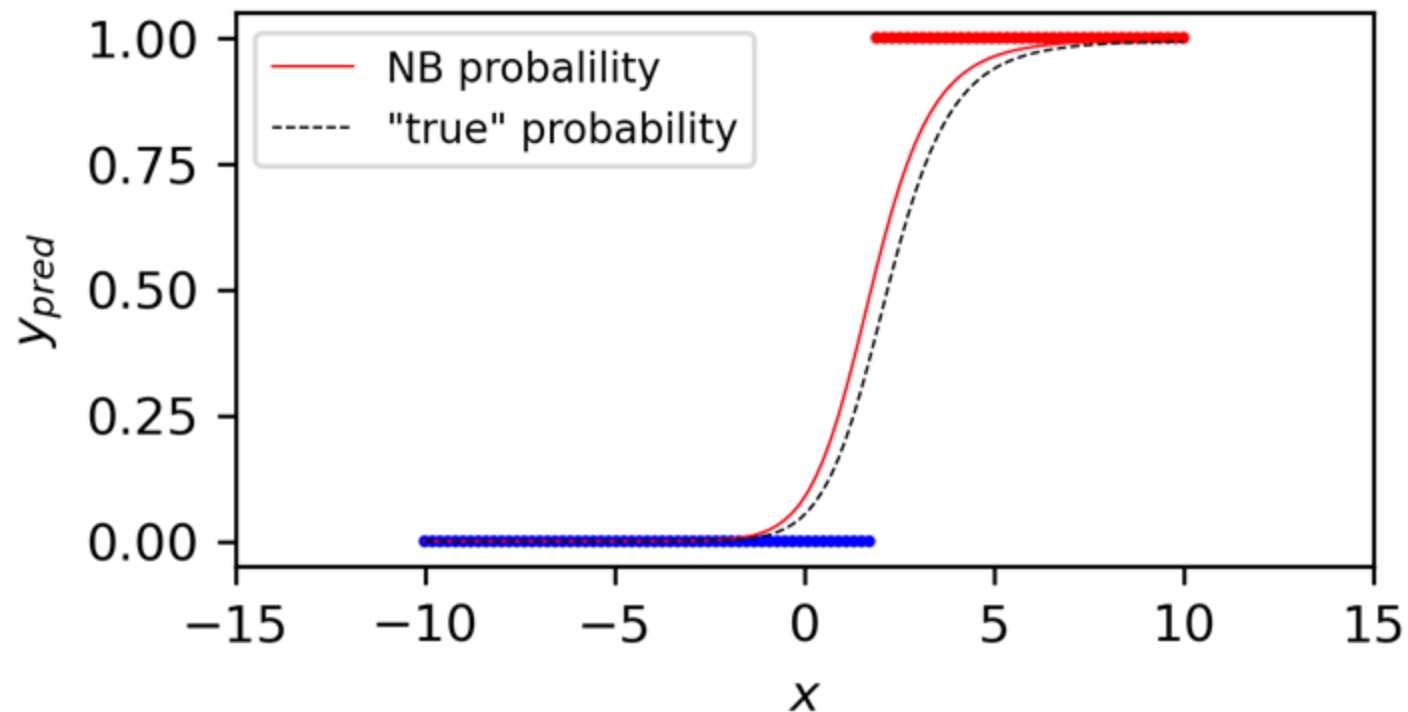
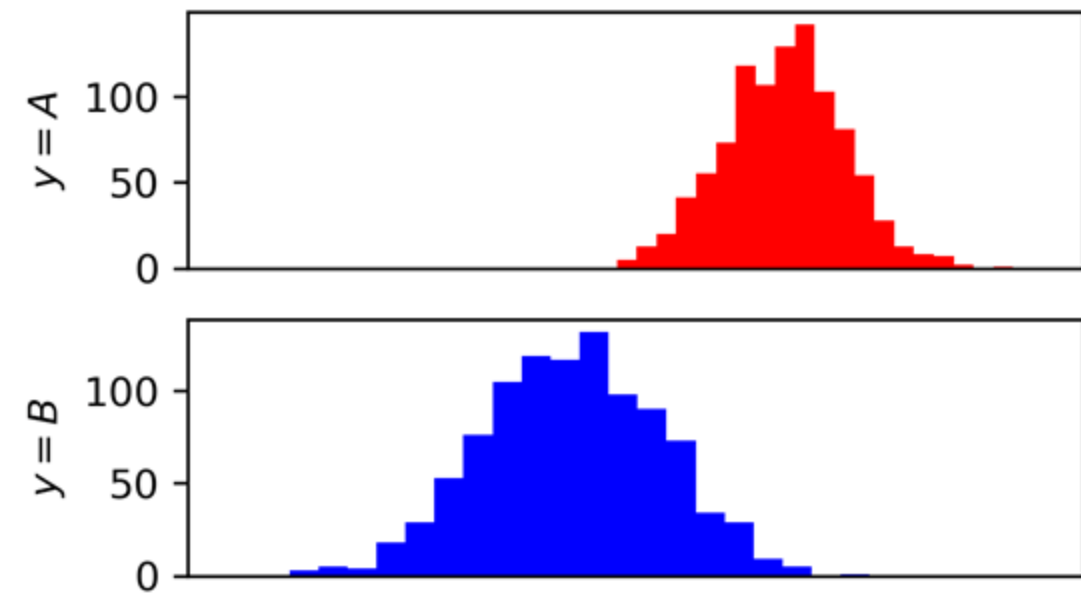
Остается оценить распределения $P(x_k | Y)$ для всех k независимо друг от друга – и можно подставлять в оценку вероятности $P(Y|X)$.

Оценка распределений $P(x_k | Y)$ может базироваться на предположении об их нормальности, - тогда нужно оценить:

- выборочное среднее (оценка параметра μ)
- выборочную дисперсию (оценка параметра σ^2).

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

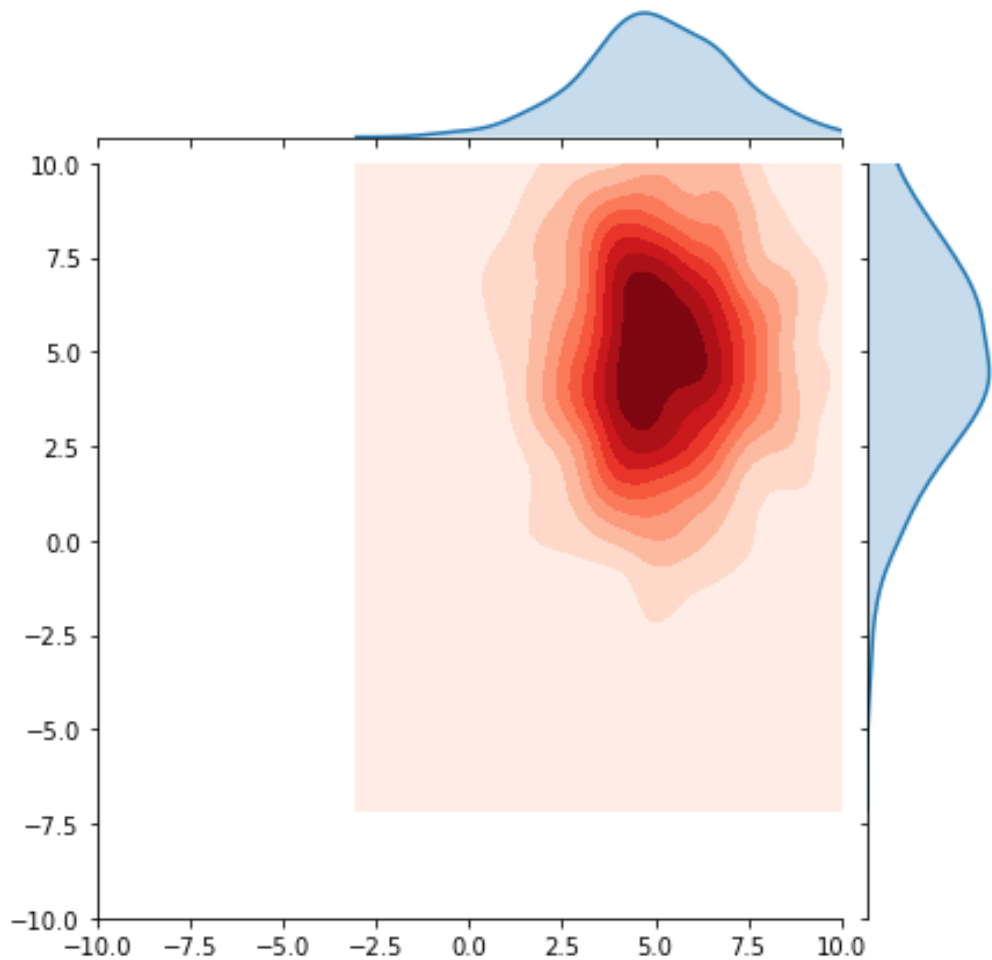
Naïve bayes classifier, «наивный байесовский классификатор»



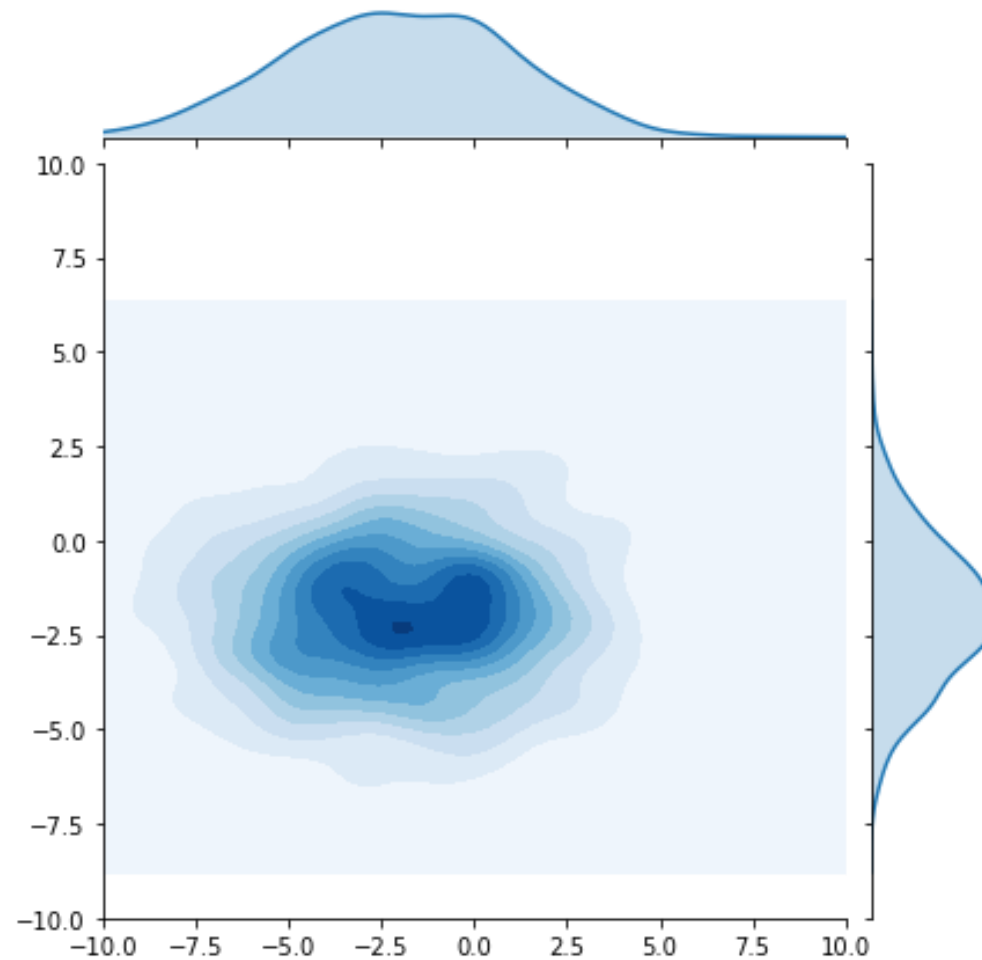
ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

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

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



$Y = A$

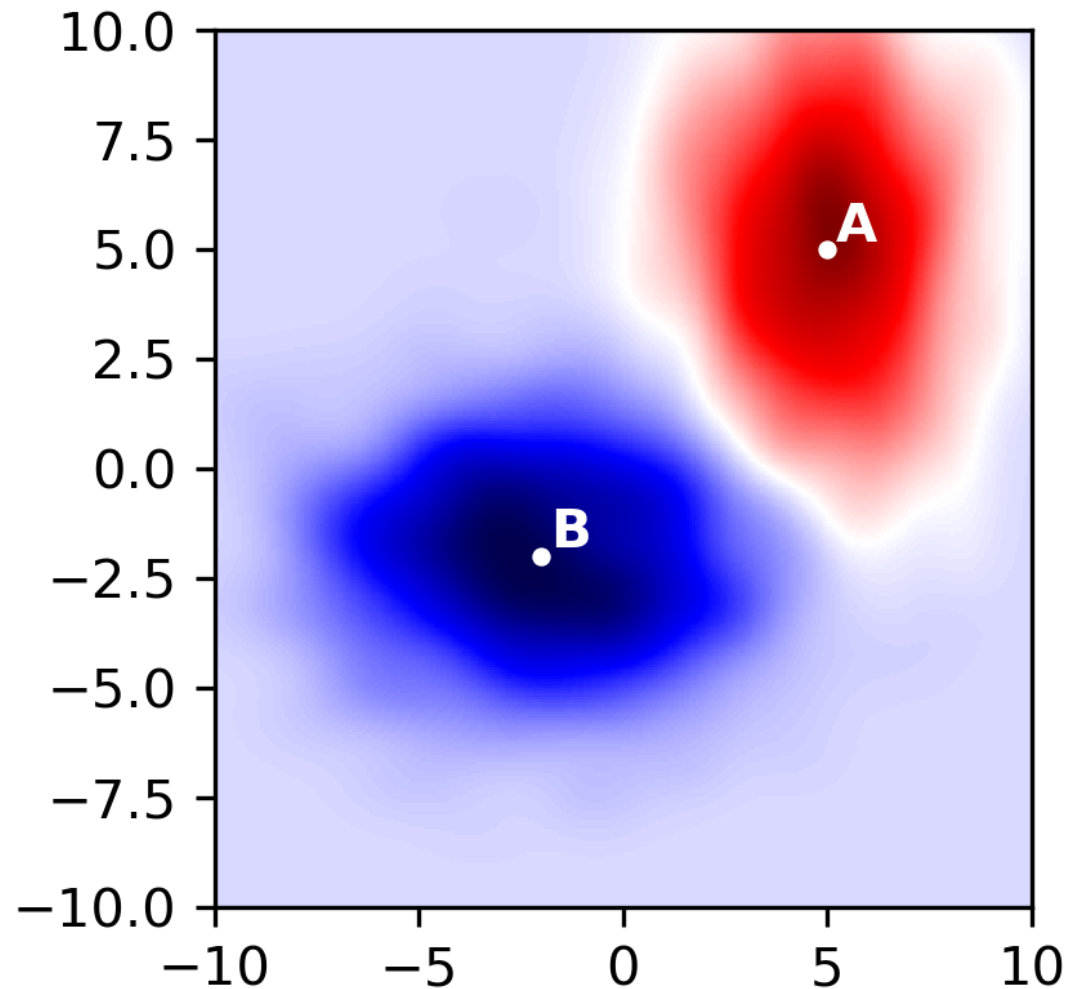


$Y = B$

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:

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

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.

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»

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

Наивное предположение об условной независимости предикторов

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

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

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

Вычисление вывода модели

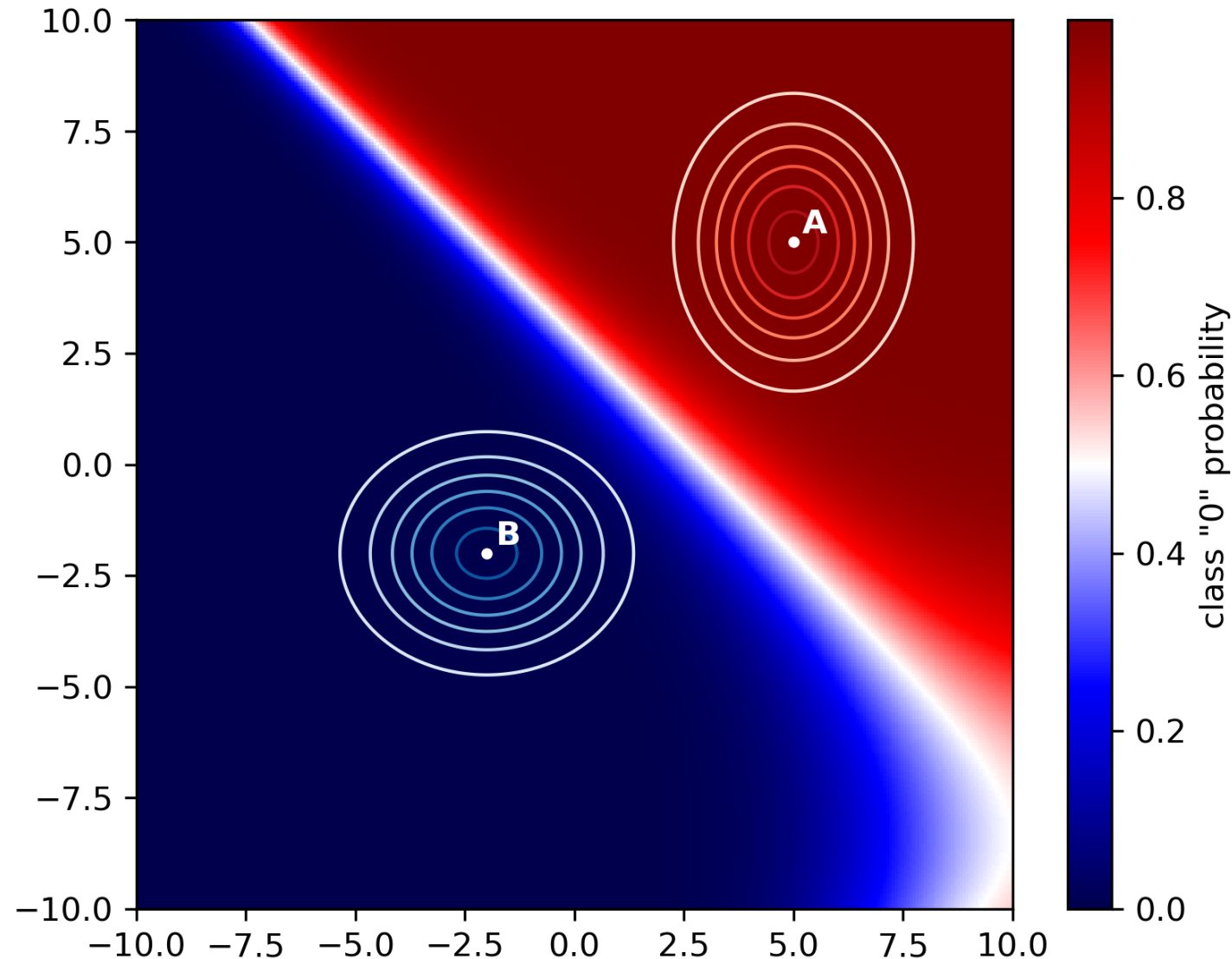
$$\hat{y}(x) = \operatorname{argmax}(P_k(x))$$

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

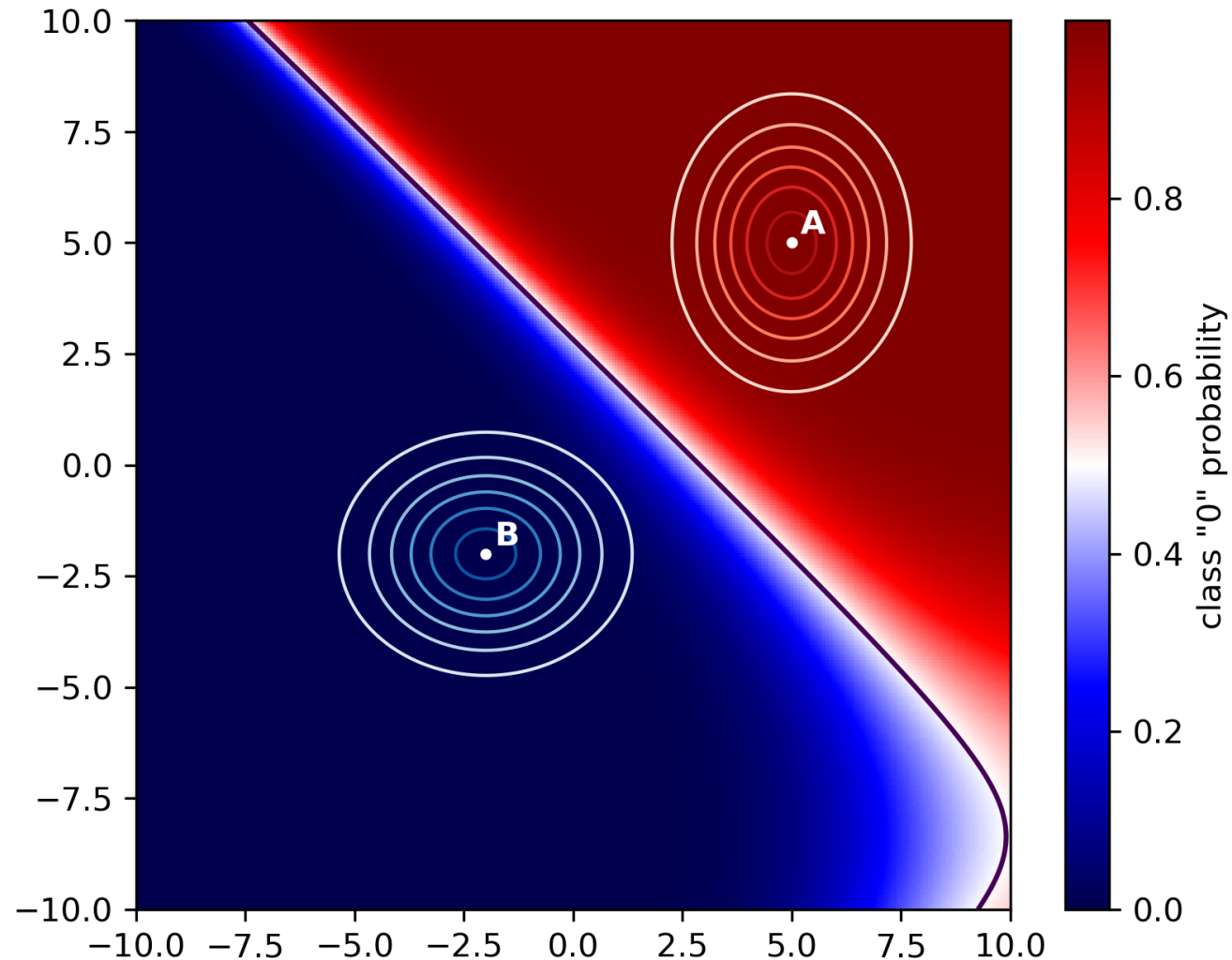
Naïve bayes classifier, «наивный байесовский классификатор»



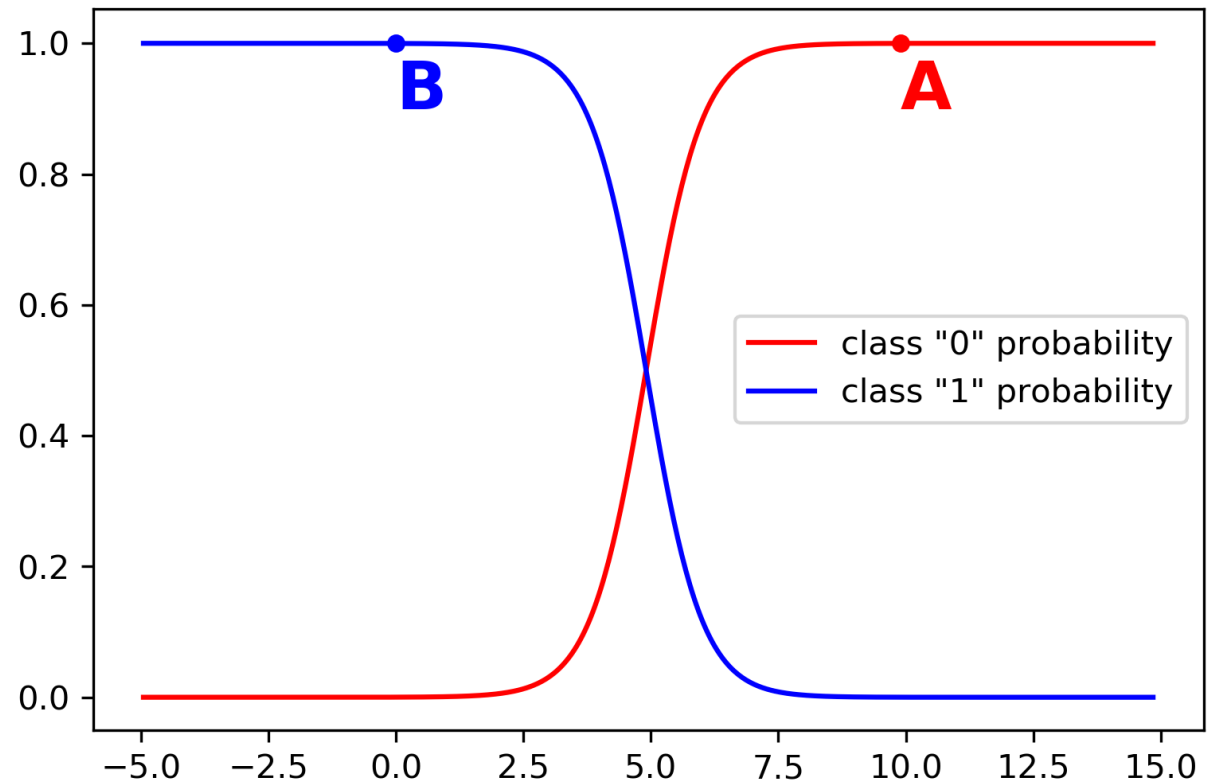
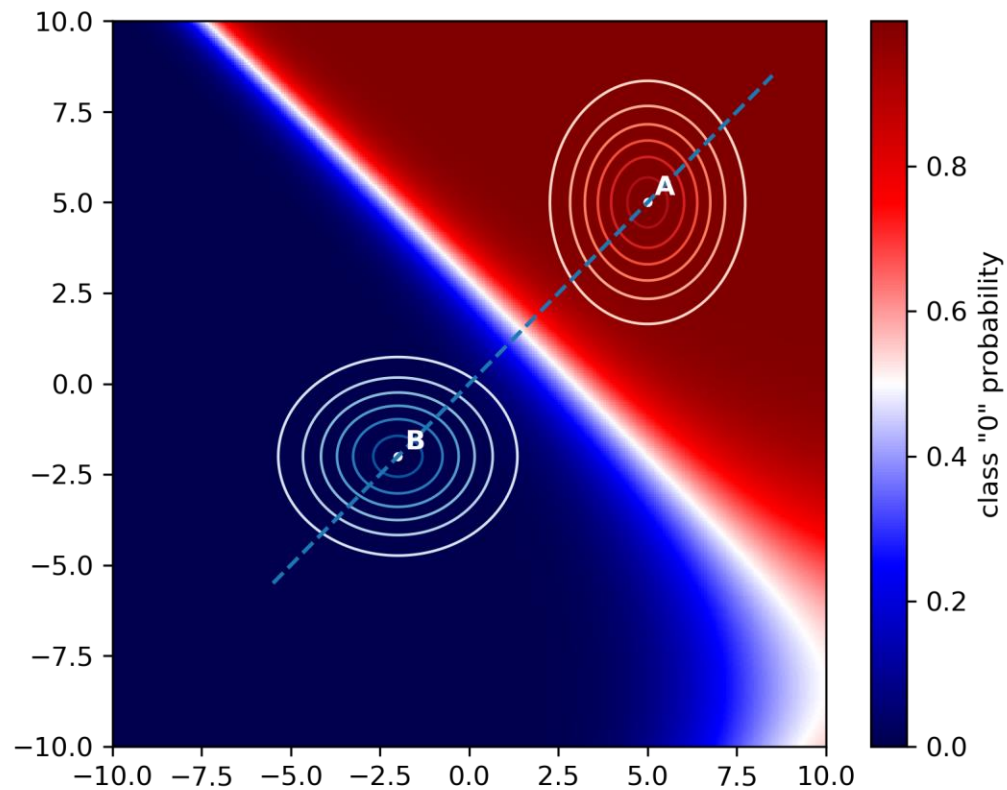
ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

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?