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

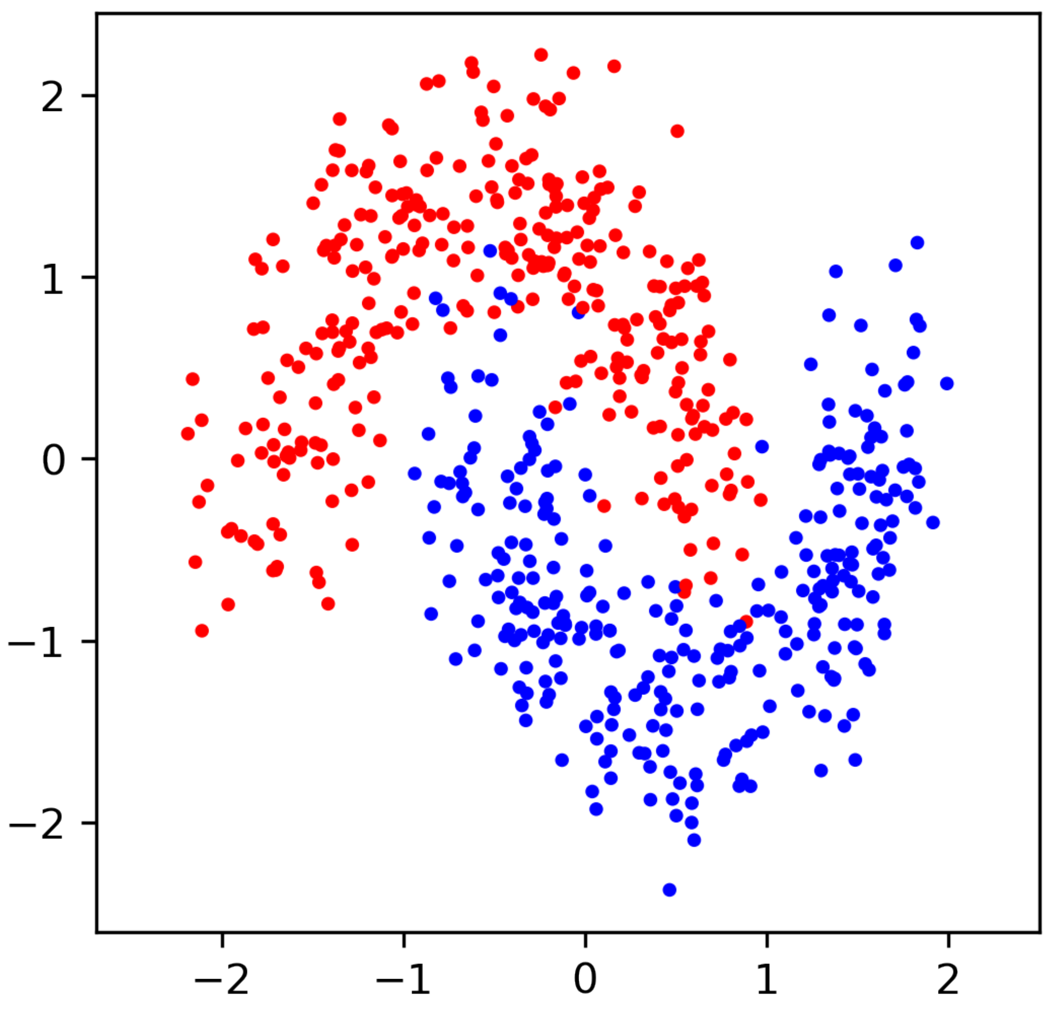
Михаил Криницкий

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)

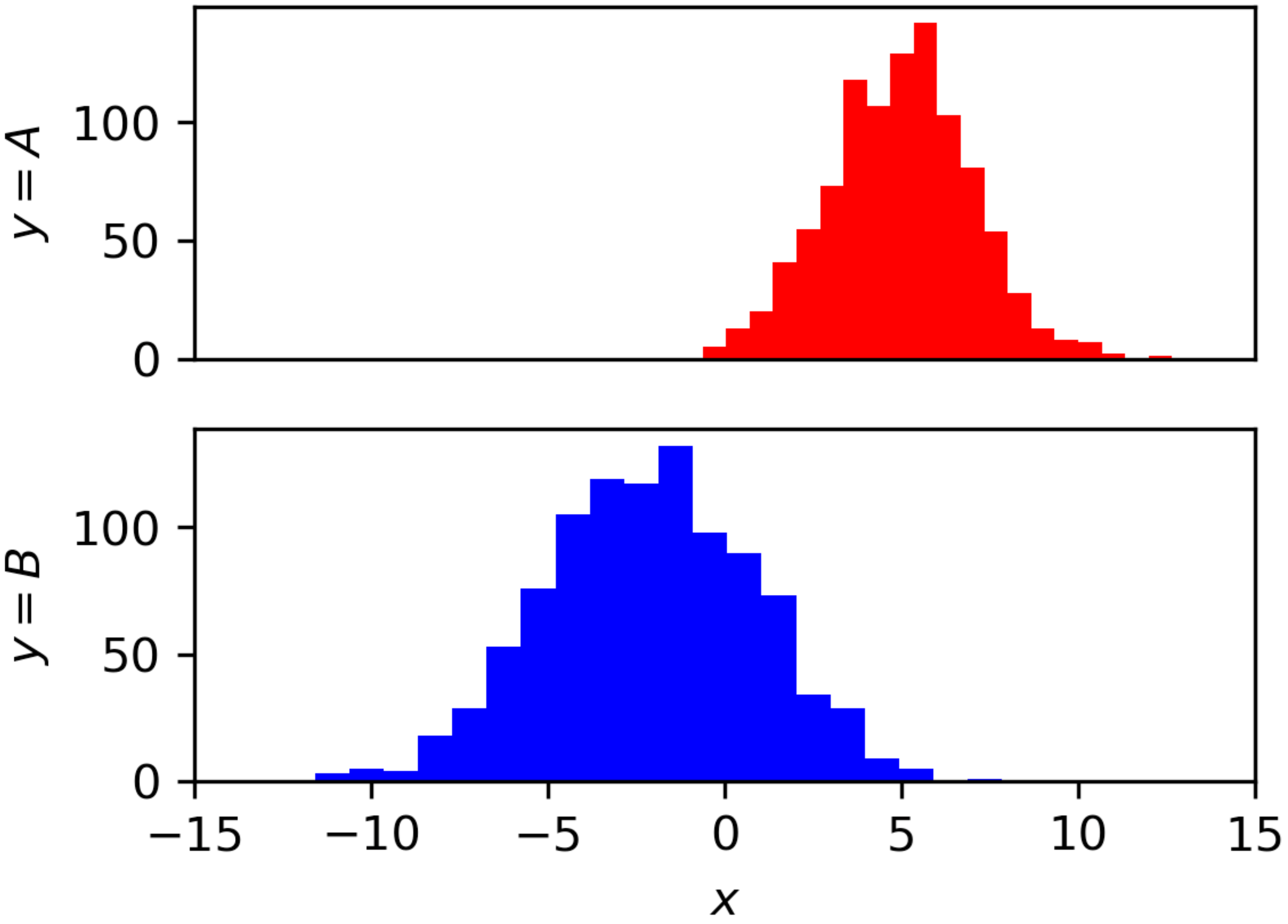
«спам / не-спам»	y – категориальная, бинарная
«мезоциклон / не-мезоциклон»	y – категориальная, бинарная
«кот / собака / лошадь»	y – категориальная, 3 класса
«0 / 1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9»	y – категориальная, 10 классов
«есть дельфин / нет дельфина»	y – категориальная, бинарная

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

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

Базируясь на этих данных, каково должно быть решение (значение $\boldsymbol{y}$) при:

$$x = -10$$
$$x = -5$$
$$x = 2$$
$$x = 5$$
$$x = 10$$
$$x = 15$$

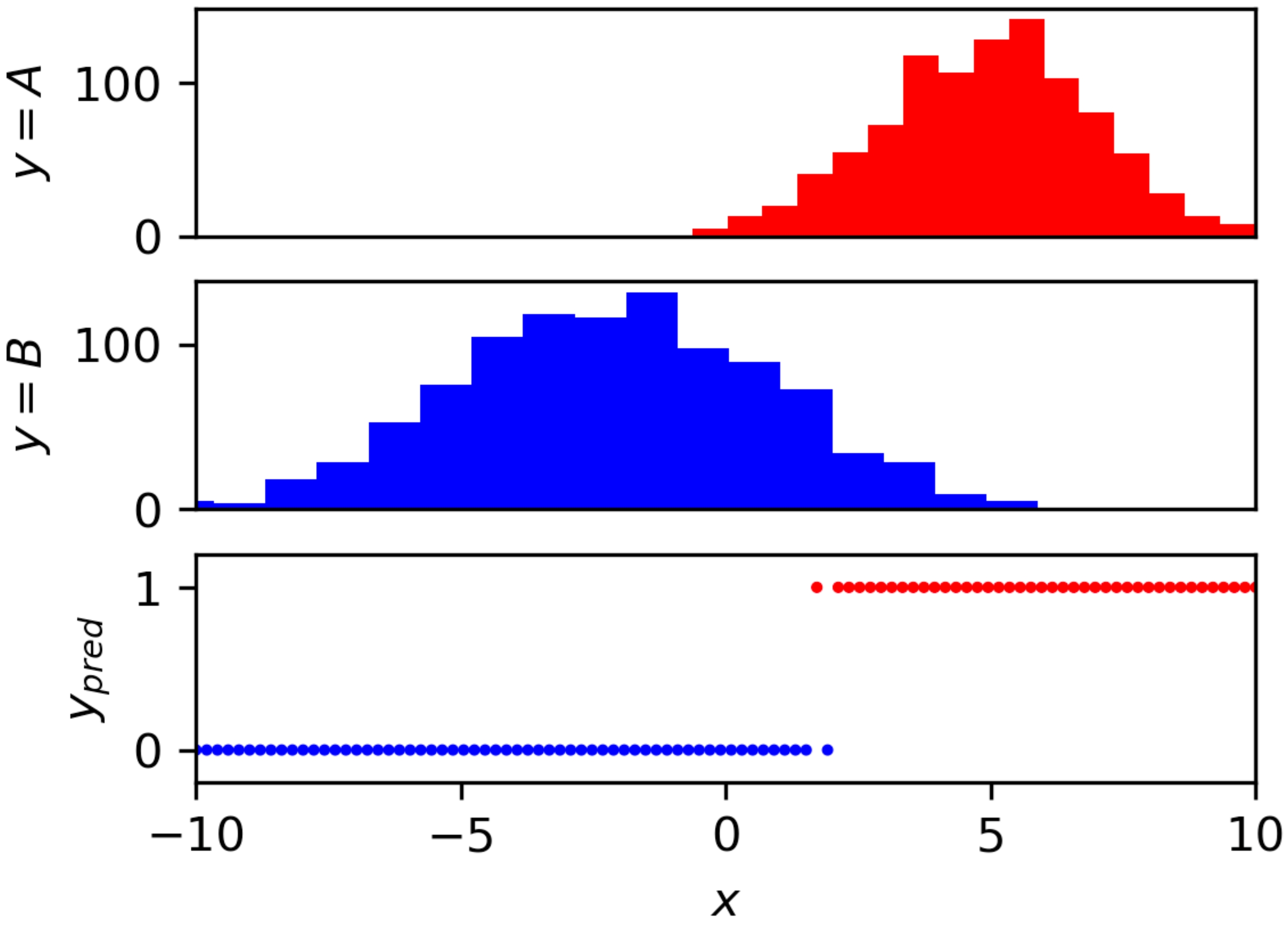


ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

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

Подход №1: **KNN** (метод $\boldsymbol{K}$ ближайших соседей)

1. выбрать $\boldsymbol{K}$ 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



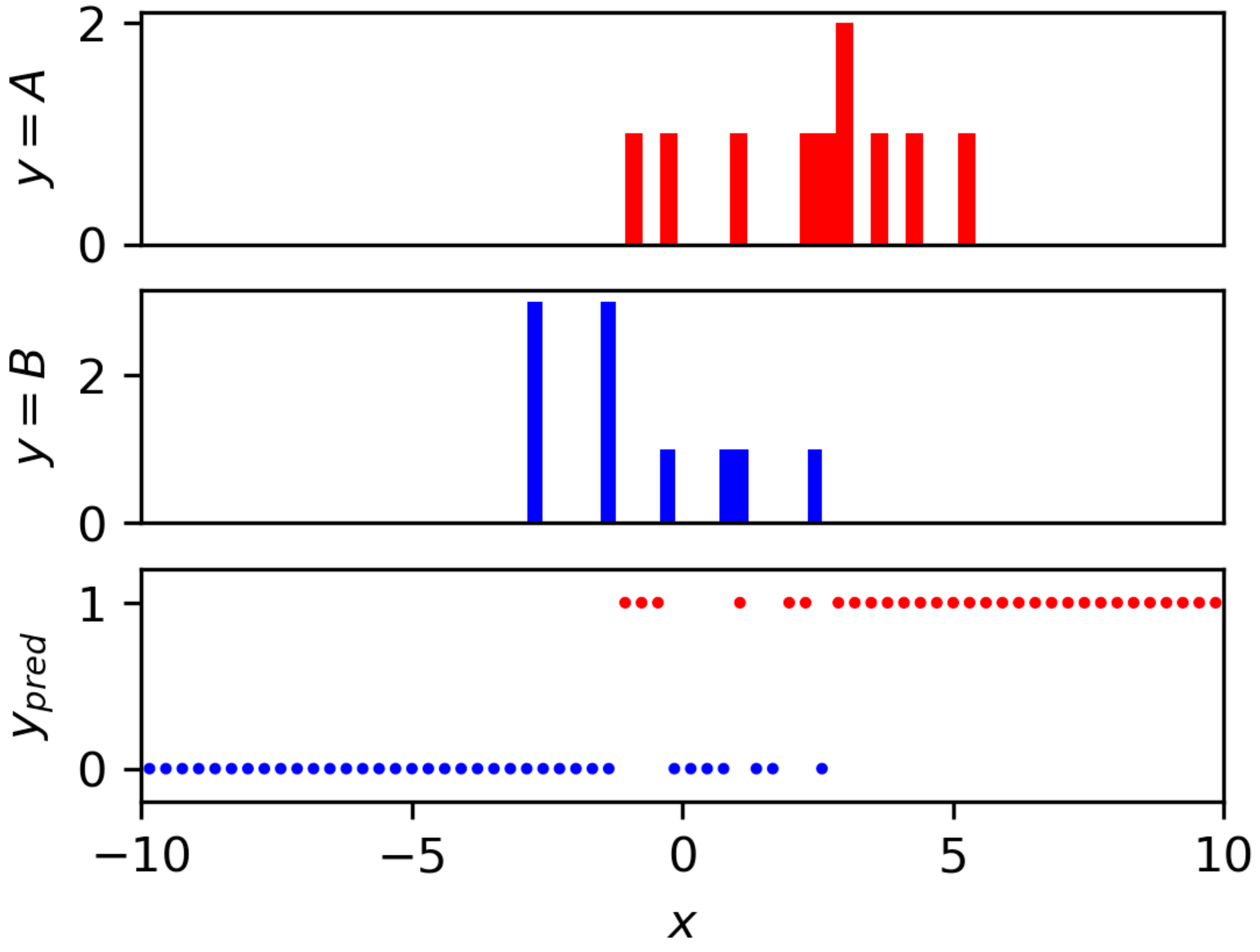
ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

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 $\boldsymbol{A}$ и $\boldsymbol{B}$ 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 $P(Y = \boldsymbol{A}|X = x_i)$ и $P(Y = \boldsymbol{B}|X = x_i)$

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

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 $P(Y = B|X = x_0)$

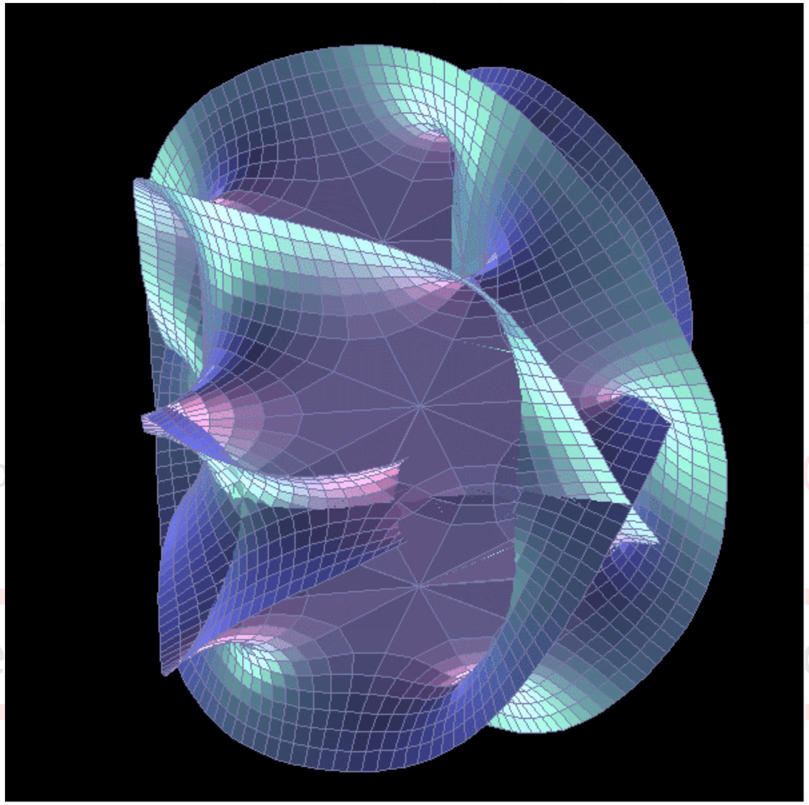
ЕСЛИ нам повезло и МЫ ЗНАЕМ* распределения X для каждого из классов $P(X|Y = \boldsymbol{A})$, $P(X|Y = \boldsymbol{B})$ etc., - то можно получить **аналитическое решение!**

И это решение будет ЛУЧШИМ из всех возможных.

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

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

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



Кстати, если нужно **принять решение** о ... енном значении x_0, помни, что $P(x_0)$ – константа, которую можно не учитыват ... $Y = B|X = x_0)$

ЕСЛИ нам повезло и МЫ ЗНАЕМ* распр ... $(X|Y = \boldsymbol{A})$, $P(X|Y = \boldsymbol{B})$ etc., - то можно получить **аналитическое решение!**

И это реше ... озможных.

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

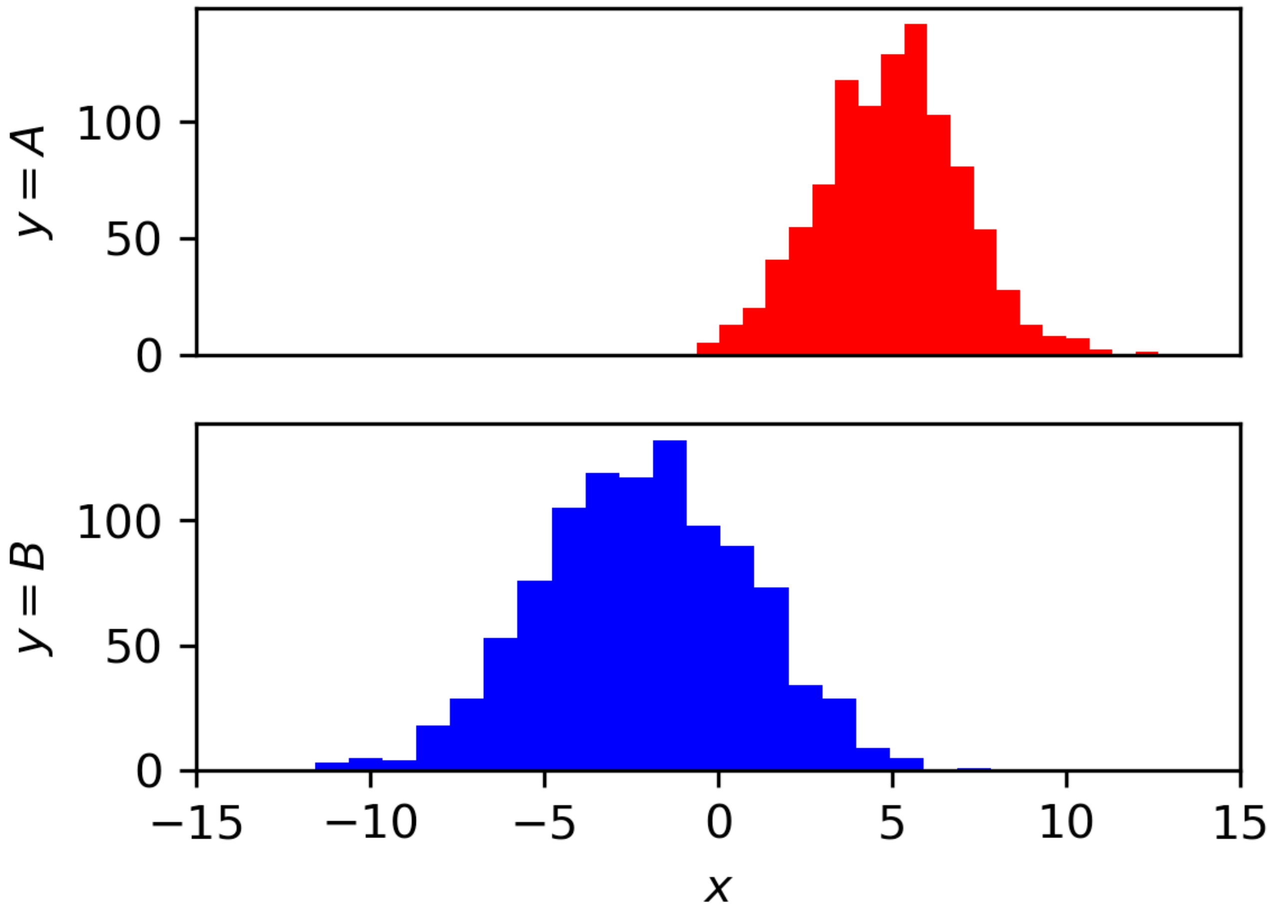
ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

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

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

объекты описываются действительным признаком $\boldsymbol{x}$
целевая переменная $\boldsymbol{y}$ – бинарная
пусть для класса $y = \boldsymbol{A}$ значения $x \sim \mathcal{N}(\mu_A, \sigma_A)$, для класса $y = \boldsymbol{B}$ значения $x \sim \mathcal{N}(\mu_B, \sigma_B)$

$$\mu_A = 5$$
$$\mu_B = -2$$
$$\sigma_A = 2$$
$$\sigma_B = 3$$



ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

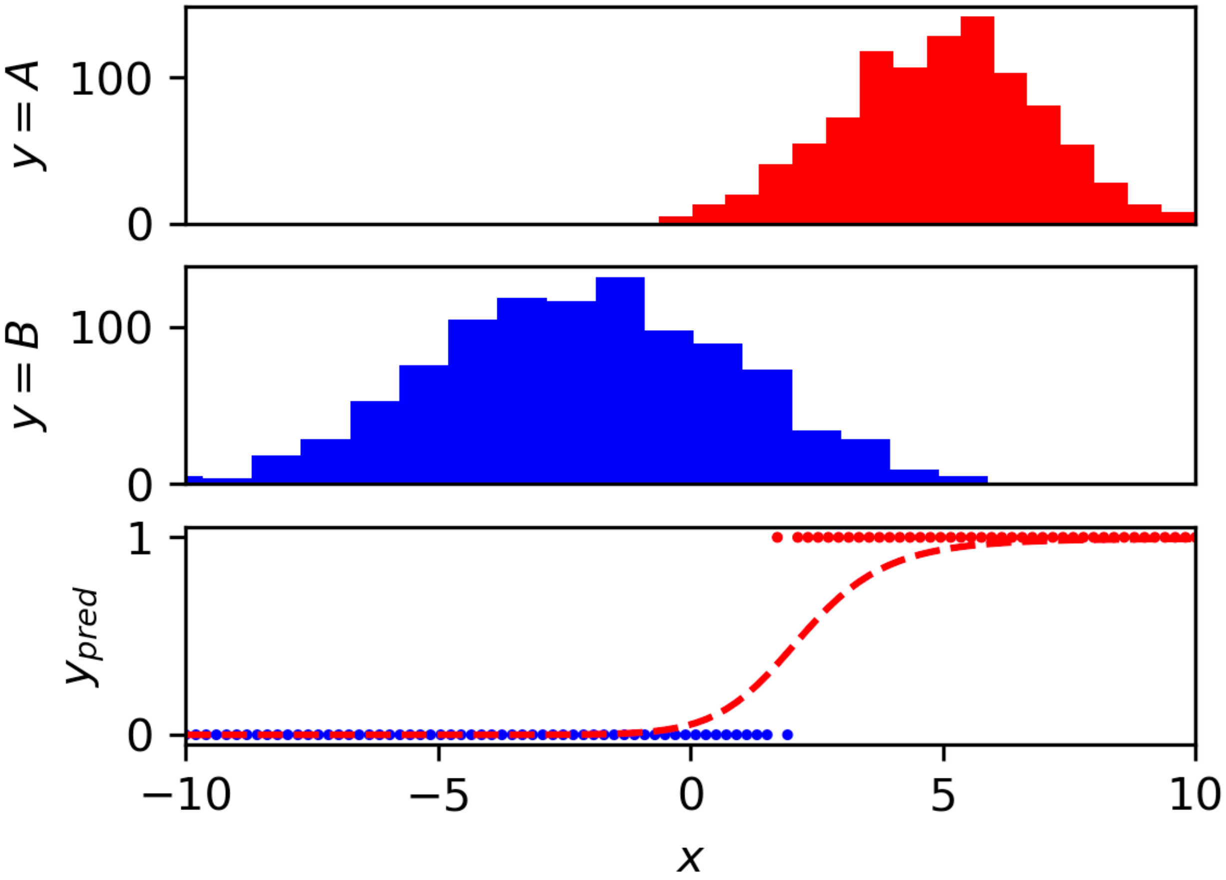
$$P(Y|X) = \frac{P(X|Y)\ P(Y)}{P(X)}$$

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

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

«Байесовский классификатор»

(не путать с «naïve bayes»)



ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

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 $P(X|Y = \boldsymbol{A})$, $P(X|Y = \boldsymbol{B})$ 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 $P(X|Y = \boldsymbol{A})$, $P(X|Y = \boldsymbol{B})$ 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 $P(X|Y = \boldsymbol{A})$, $P(X|Y = \boldsymbol{B})$ 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 $P(X = x)$

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

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

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Naïve bayes classifier, «наивный байесовский классификатор»

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

можно забыть про $P(X = x)$

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

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

наивное предположение: переменные x_i условно независимы:

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

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

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Naïve bayes classifier, «наивный байесовский классификатор»

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

можно забыть про $P(X=x)$

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

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

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 $\boldsymbol{P(x_k|Y)}$ для всех $\boldsymbol{k}$ 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 $P(X = x)$

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

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

наивное предположение: переменные x_i условно независимы:

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

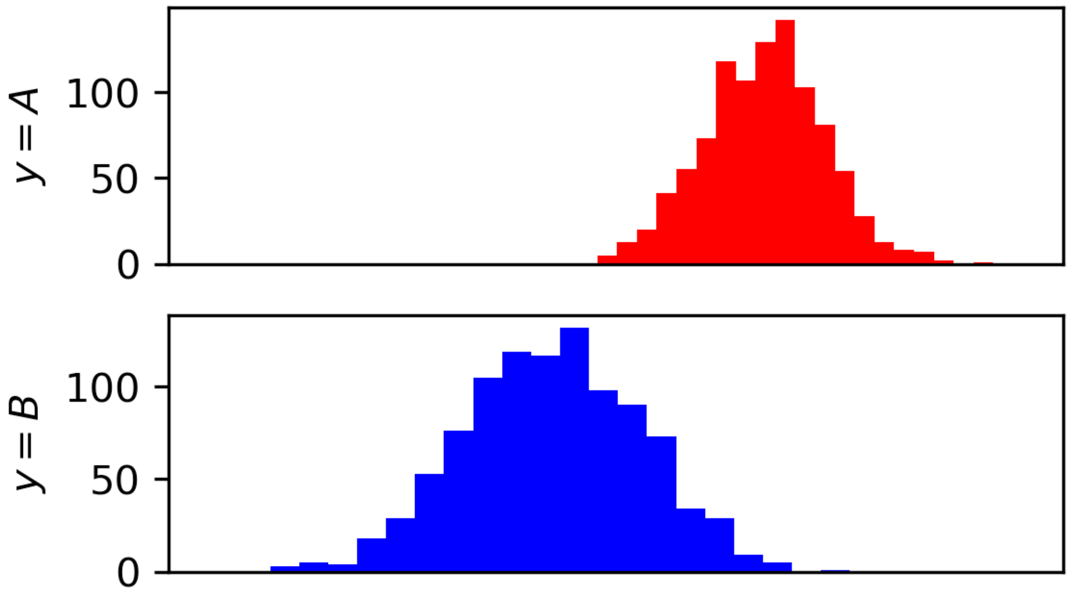
Остается оценить распределения $\boldsymbol{P(x_k|Y)}$ для всех $\boldsymbol{k}$ независимо друг от друга – и можно подставлять в оценку вероятности $P(Y|X)$.

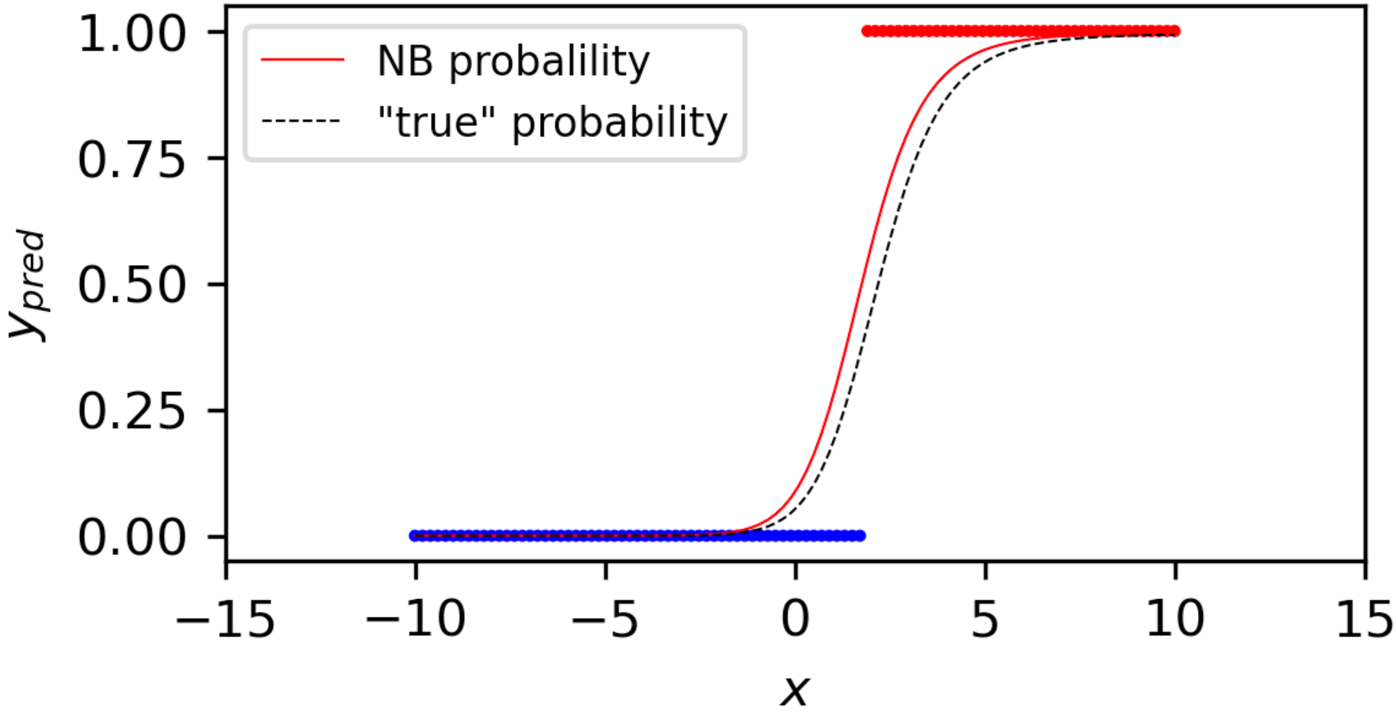
Оценка распределений $\boldsymbol{P(x_k|Y)}$ может базироваться на предположении об их нормальности, - тогда нужно оценить:

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

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Naïve bayes classifier, «наивный байесовский классификатор»

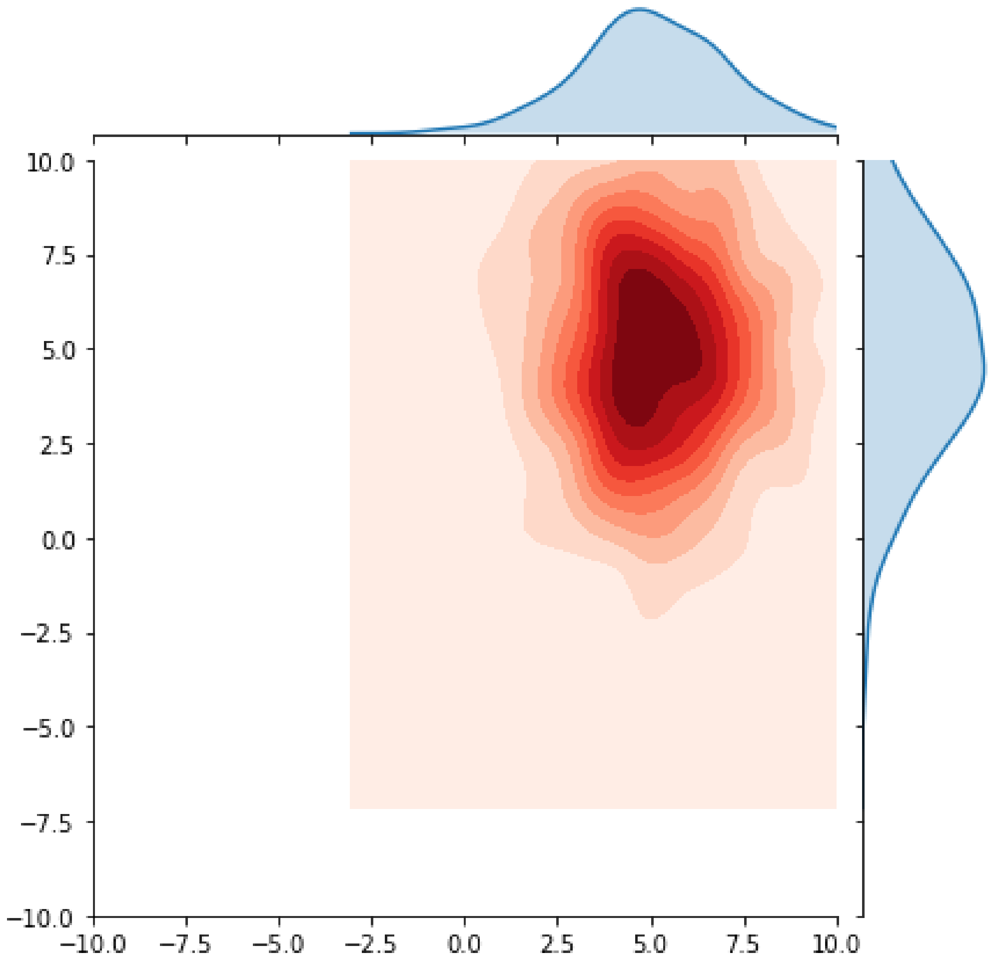




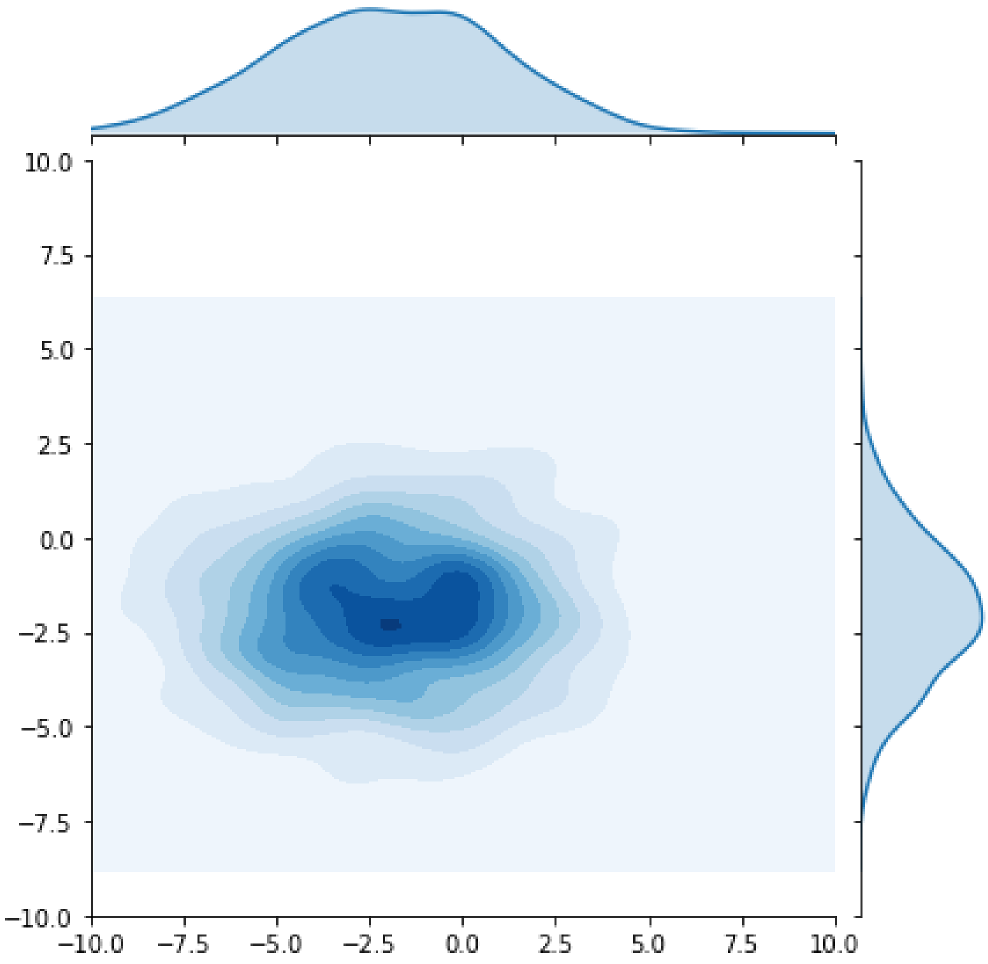
ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

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

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



$Y = A$



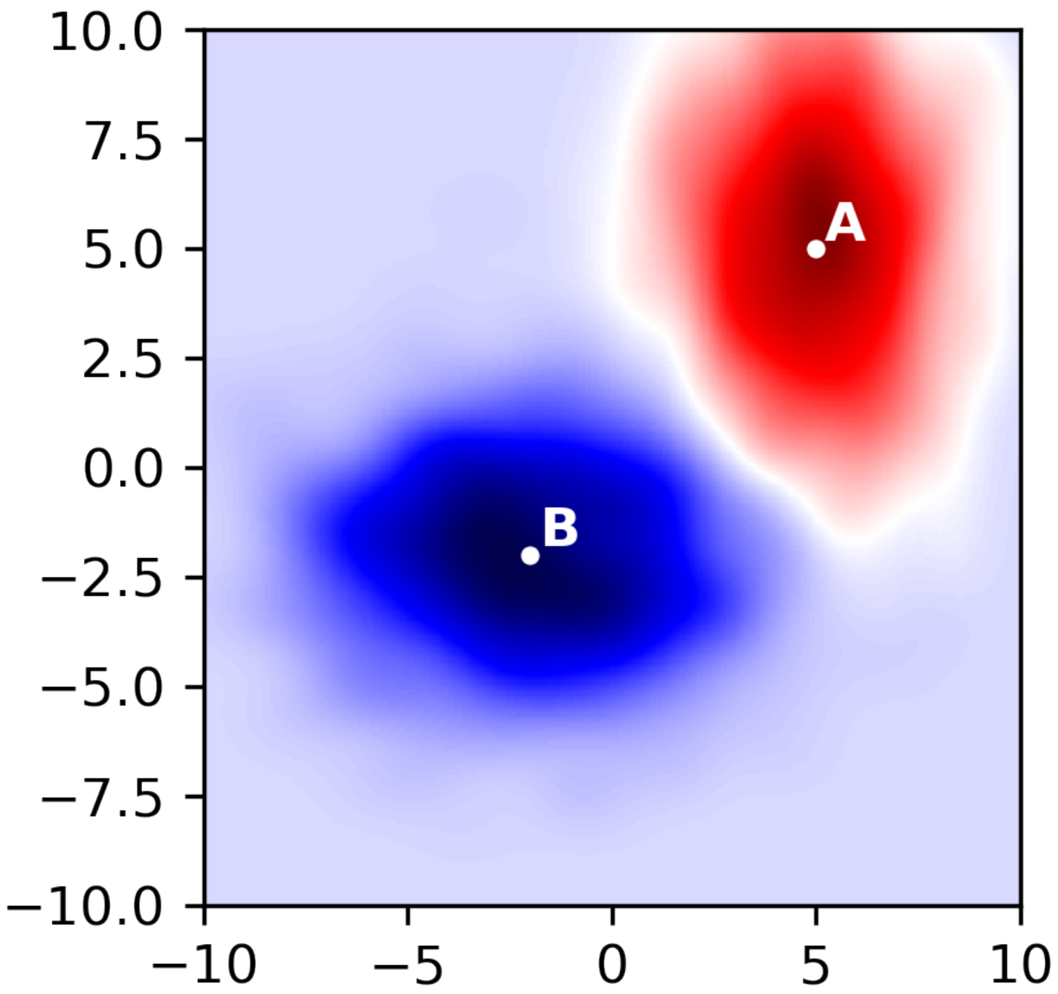
$Y = B$

previously on ML4ES

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Цель: оценить **вероятность** классов $\boldsymbol{A}$ (класс “**1**”) и $\boldsymbol{B}$ (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 $P(X = x)$

$$P(X|Y) = P\left(x_1, x_2, x_3 \ldots x_p \middle| Y\right) = P\left(x_2, x_3 \ldots x_p \middle| Y, x_1\right) * P(x_1|Y) = \cdots$$

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

наивное предположение: переменные x_i условно независимы:

$$P\left(x_p \middle| Y, x_1, x_2, \ldots, x_{p-1}\right) = P\left(x_p \middle| Y\right)$$

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 $P\left(x_p \middle| Y\right)$ - **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 $P(X=x)$

$$P(X|Y) = P\left(x_1, x_2, x_3 \ldots x_p \middle| Y\right) = P\left(x_2, x_3 \ldots x_p \middle| Y, x_1\right) * P(x_1|Y) = \cdots$$

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

наивное предположение: переменные x_i условно независимы:

$$P\left(x_p \middle| Y, x_1, x_2, \ldots, x_{p-1}\right) = P\left(x_p \middle| Y\right)$$

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 $P\left(x_p \middle| Y\right)$ - **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

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

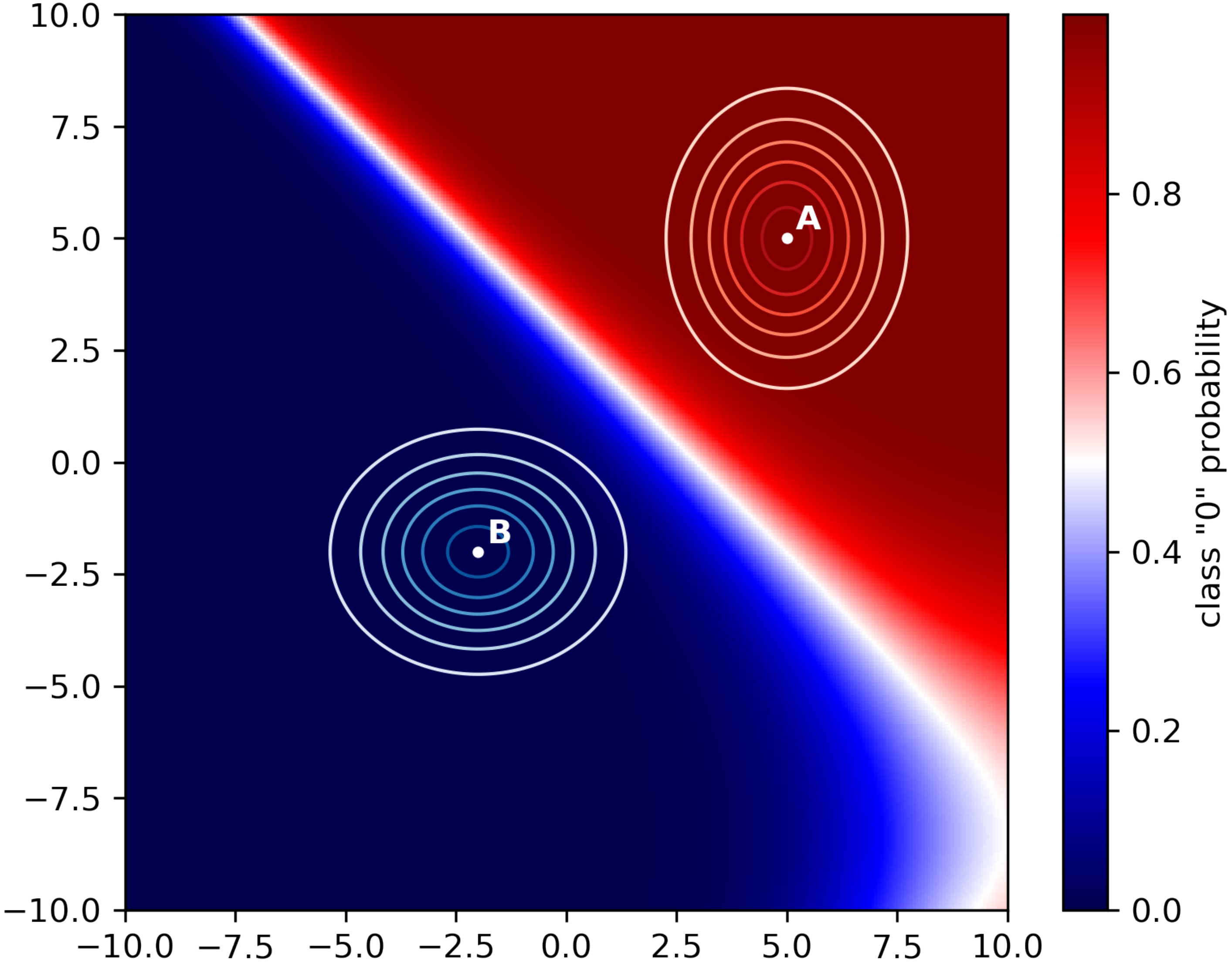
Нормировка вероятностей (если нужны оценки вероятностей)

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

code demonstration

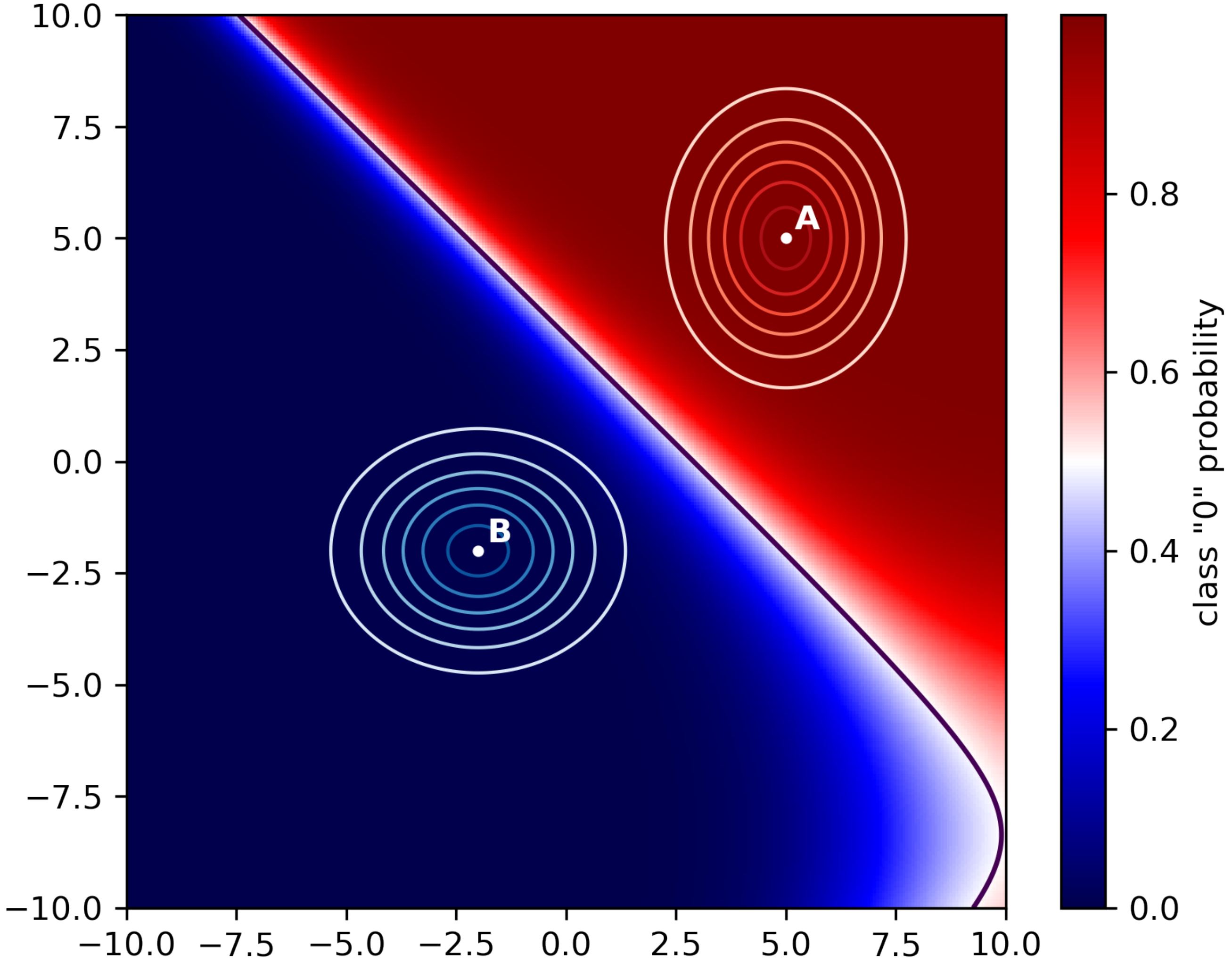
ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Naïve bayes classifier, «наивный байесовский классификатор»

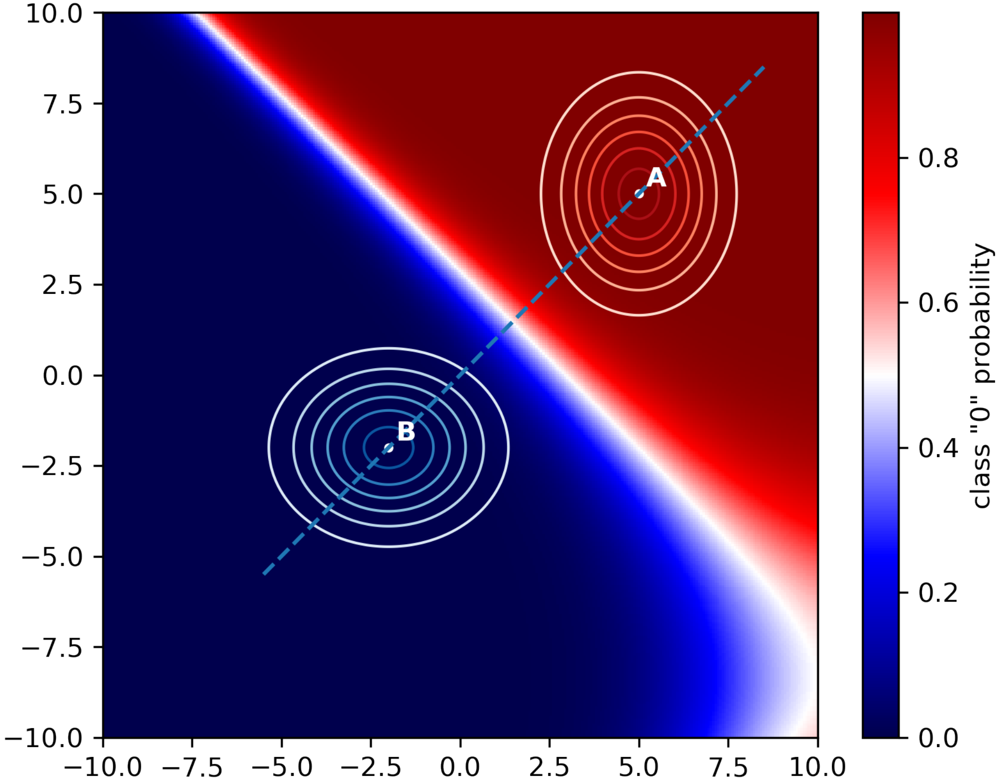


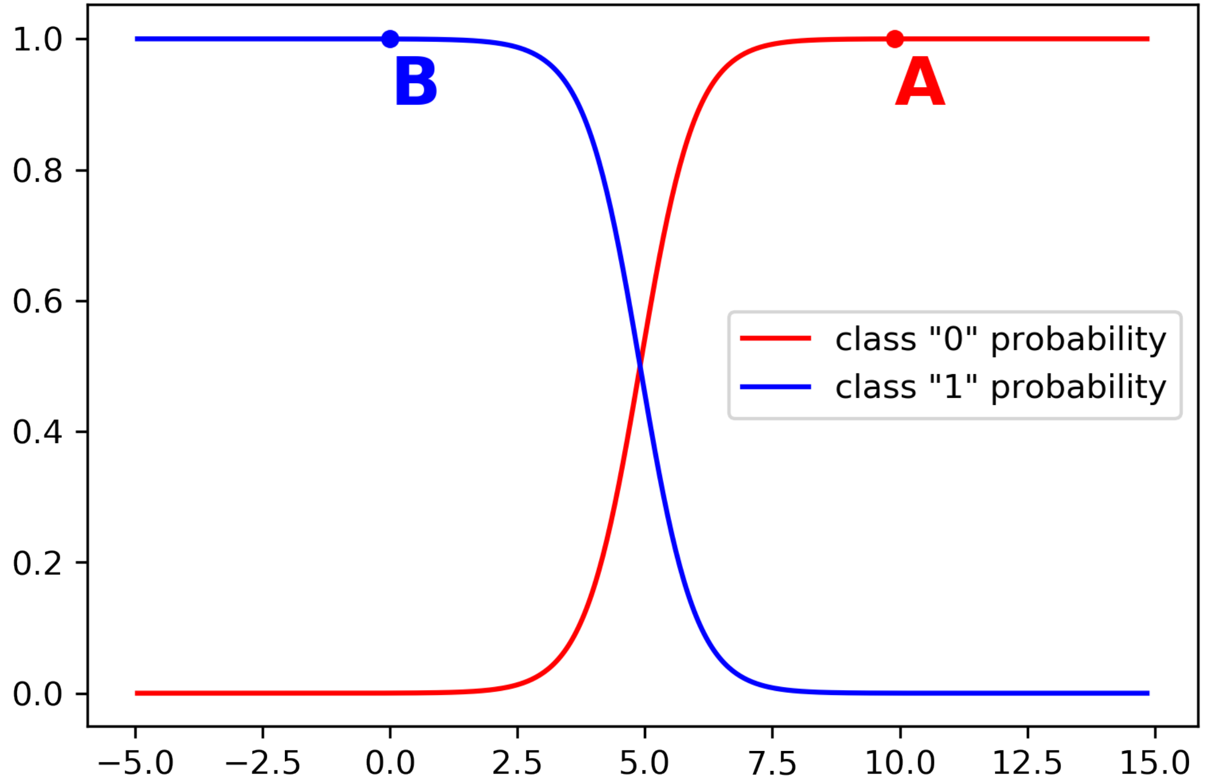
ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

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?

<u>ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации</u>

Логистическая регрессия («logistic regression»)

Линейная регрессия: набор предположений “LINE”

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

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

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

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

Логистическая регрессия:

(1) Предполагаем, что переменная Y распределена согласно распр-ю Бернулли с параметром p (зависящий от x_i);
(2) Предполагаем, что отношение log-вероятностей классов «1» и «0» соотносятся как линейная функция:

$$f(\theta, x) = \theta \cdot x$$

$$Y \sim \mathcal{B}\big(p(\theta, x)\big), \qquad \log \frac{p(\theta,x)}{1-p(\theta,x)} = \theta \cdot x$$

$$p(\theta, x) = \frac{1}{1 + e^{-\theta\cdot x}}$$

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

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Логистическая регрессия («logistic regression»)

$\log\frac{p(\theta,x)}{1-p(\theta,x)} = \theta\cdot x,\qquad p(\theta,x) = \frac{1}{1+e^{-\theta\cdot x}}$

Новое обозначение (для краткости): $p_i \equiv p(\theta, x_i)$

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

Бинарная перекрестная энтропия (отрицательная)
Negative binary cross-entropy

$$\ell(\mathcal{T},\theta) = \log(L(\mathcal{T},\theta)) = \sum_i \log\left(p_i^{y_i}\right) + \sum_i \log\left((1-p_i)^{(1-y_i)}\right) =$$

$$= \sum_i (y_i * \log p_i + (1-y_i) * \log(1-p_i)) = \sum_i \log(1-p_i) + \sum_i y_i * \log\frac{p_i}{1-p_i} =^{(*)}$$

$$\left\{p_i = \frac{1}{1+e^{-\theta\cdot x_i}}\ :\ (1-p_i) = \frac{1}{1+e^{\theta\cdot x_i}}\right\}$$

$$\left\{\log\frac{p_i}{1-p_i} = \theta\cdot x_i\right\}$$

$$(*) = -\sum_i \log\left(1+e^{\theta\cdot x_i}\right) + \sum_i y_i * \theta\cdot x_i$$

$$\frac{\partial\ell(\mathcal{T},\theta)}{\partial\theta} = \sum_i (y_i - p(\theta,x_i))x_i$$

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Логистическая регрессия («logistic regression»)

Линейная регрессия:

набор предположений “LINE”

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

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

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

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

Логистическая регрессия:

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

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

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

$$\ell(\mathcal{T}, \theta) = -\sum_{i} \log\big(1 + e^{\theta \cdot x_i}\big) + \sum_{i} y_i * \theta \cdot x_i$$

$$\theta^* = \underset{\Theta}{\operatorname{argmax}}\big(\ell(\mathcal{T}, \theta)\big) = \underset{\Theta}{\operatorname{argmin}} \left(\sum_{i} \big(\log\big(1 + e^{\theta \cdot x_i}\big) - y_i * \theta \cdot x_i\big)\right)$$

$$\underset{\Theta}{\operatorname{argmax}} \sum_{i} \big(y_i * \log p_i + (1 - y_i) * \log(1 - p_i)\big)$$

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Логистическая регрессия («logistic regression»)

code demonstration

ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Разделяющая поверхность
(на примере результатов логистической регрессии)