



# 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 (продолжение)

Михаил Криницкий

К.Т.Н., С.Н.С.

Институт океанологии РАН им. П.П. Ширшова

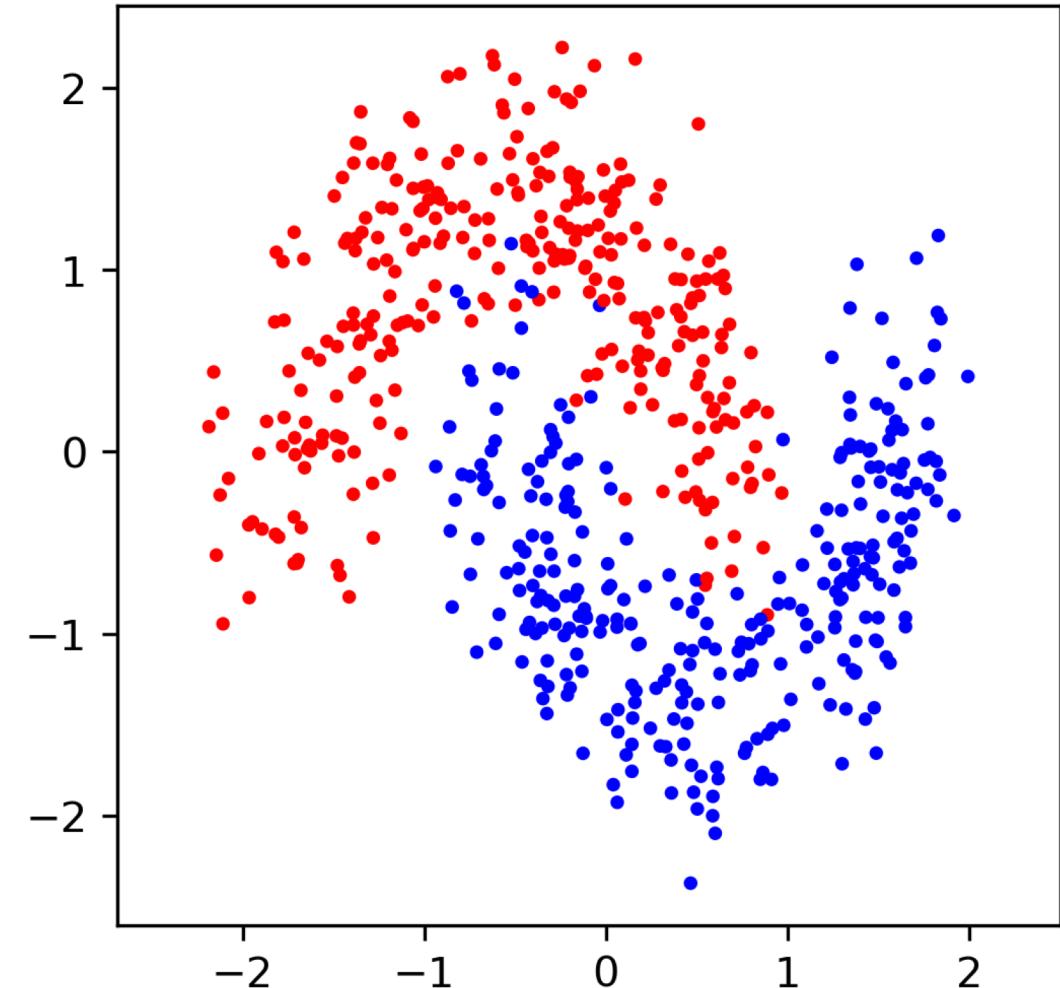
Лаборатория взаимодействия океана и атмосферы и  
мониторинга климатических изменений (ЛВОАМКИ)

# КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

## задача в терминах машинного обучения

- «Обучение с учителем»
  - восстановление регрессии
  - классификация

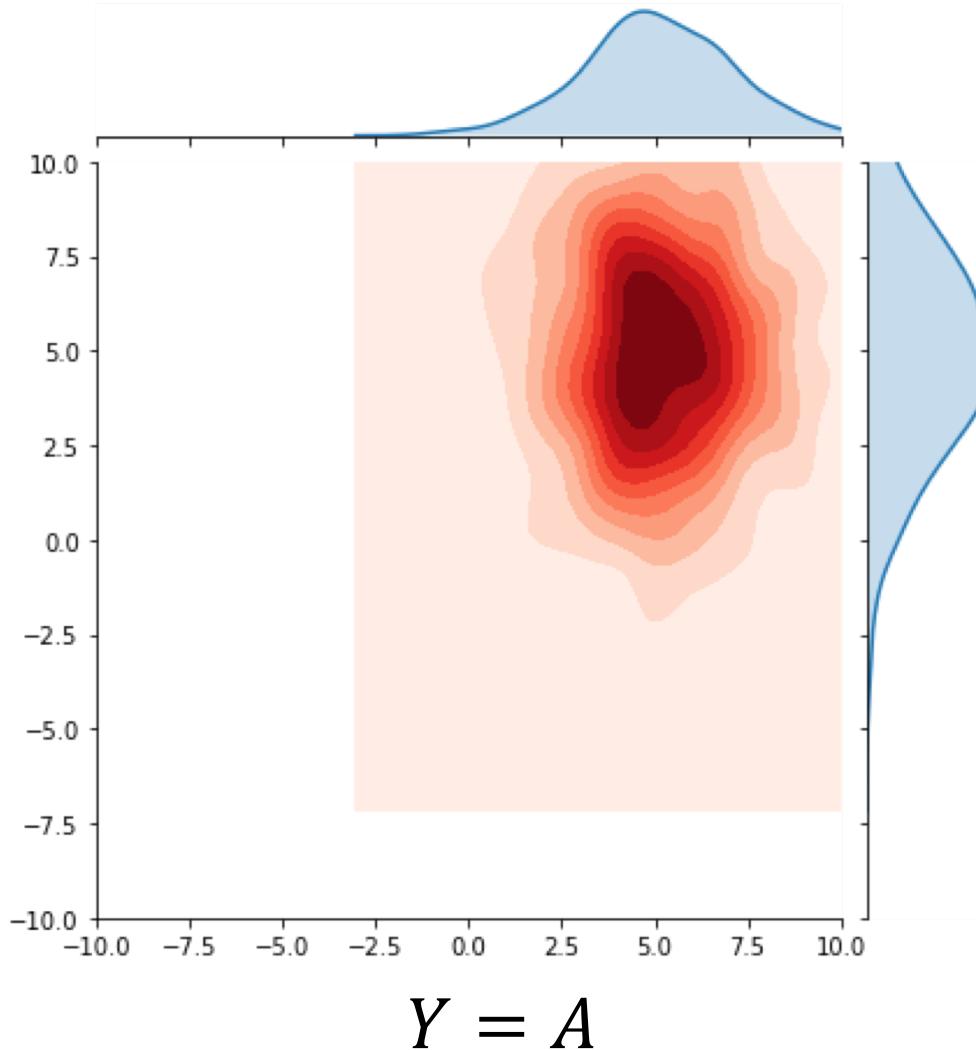
**ЧТО Я ХОЧУ?** – метку класса  
**«красный или синий?»**  
(бинарная классификация)



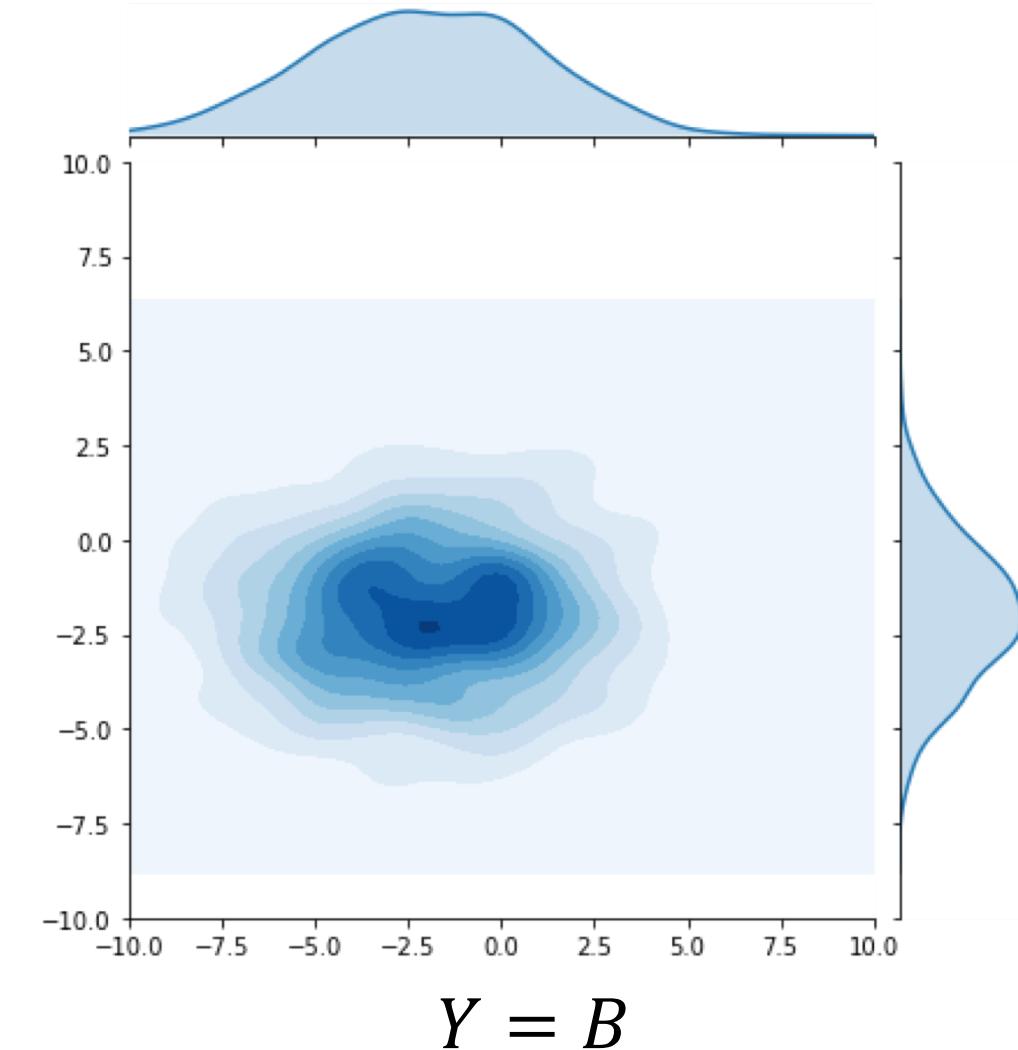
# ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

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

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



$Y = A$

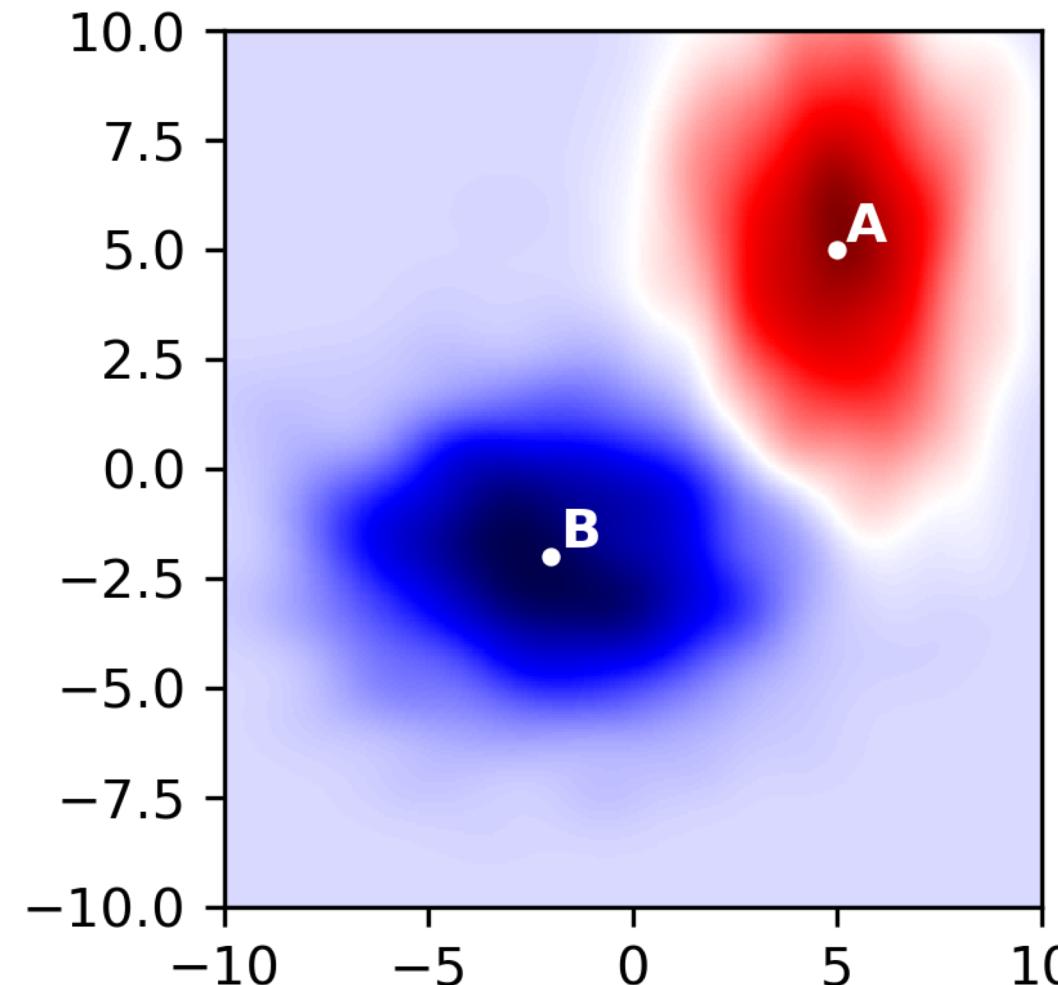


$Y = B$

# 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»

$$P(Y|X) \propto P(X|Y) \cdot P(Y)$$
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  $\mu$ ) и выборочную дисперсию (оценка параметра  $\sigma^2$ ).

### NOTE!

**LDA** (Linear Discriminant Analysis, метод линейного дискриминантного анализа) – то же самое, но в предположении, что (a) распределение  $P(x_p | Y)$  - нормальное; (б) для всех компонент  $x_i$  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  $\mu$ ) и выборочную дисперсию (оценка параметра  $\sigma^2$ ).

### 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  $\sigma^2$  становится матрицей ковариаций  $\Sigma^2$ ). NB с нормальным распределением  $P(x_p | Y)$  - это QDA с диагональной  $\Sigma^2$

## 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

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


---

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

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

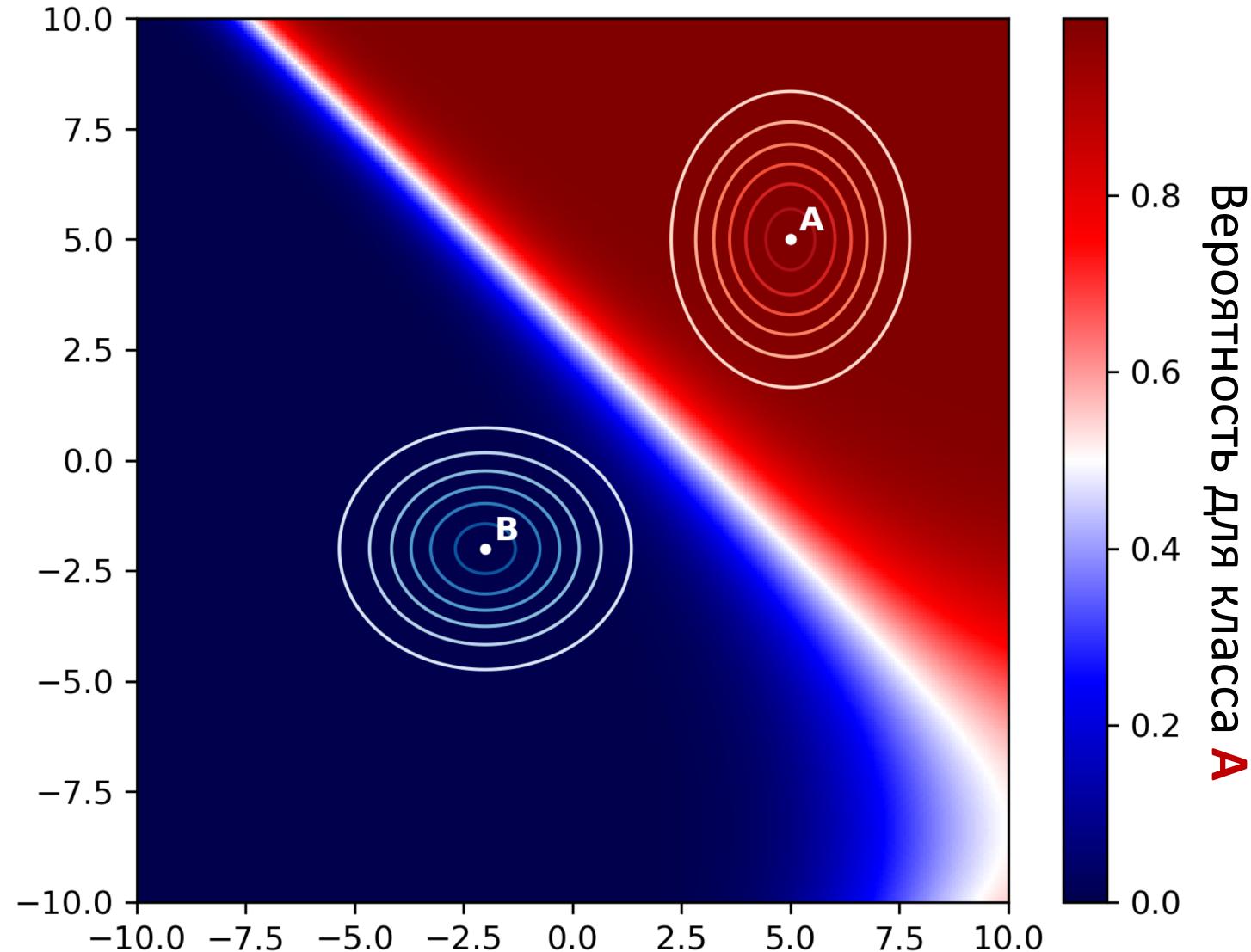
Нормировка вероятностей (если нужны оценки вероятностей)

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


---

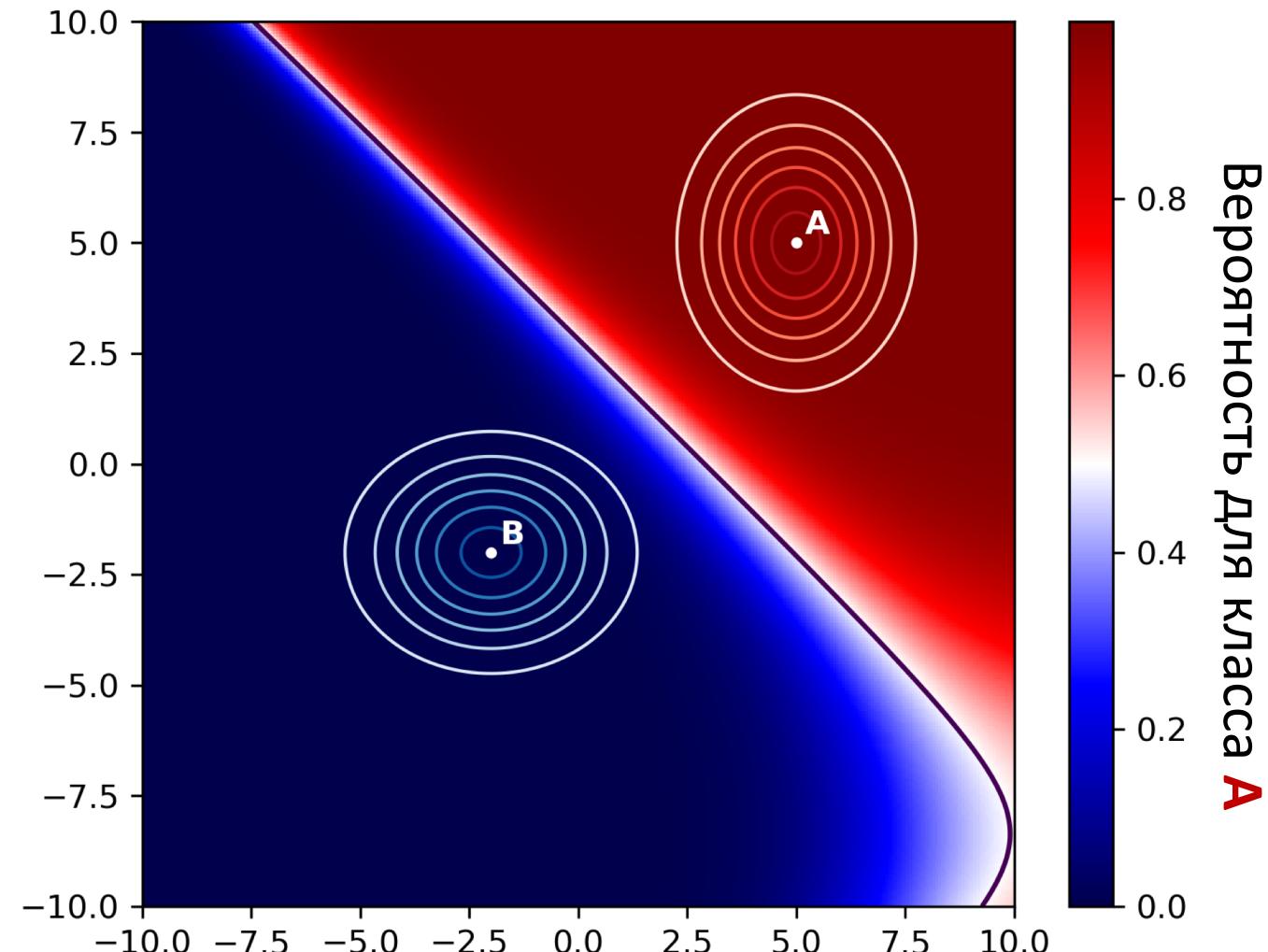
## ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Naïve bayes classifier, «наивный байесовский классификатор»

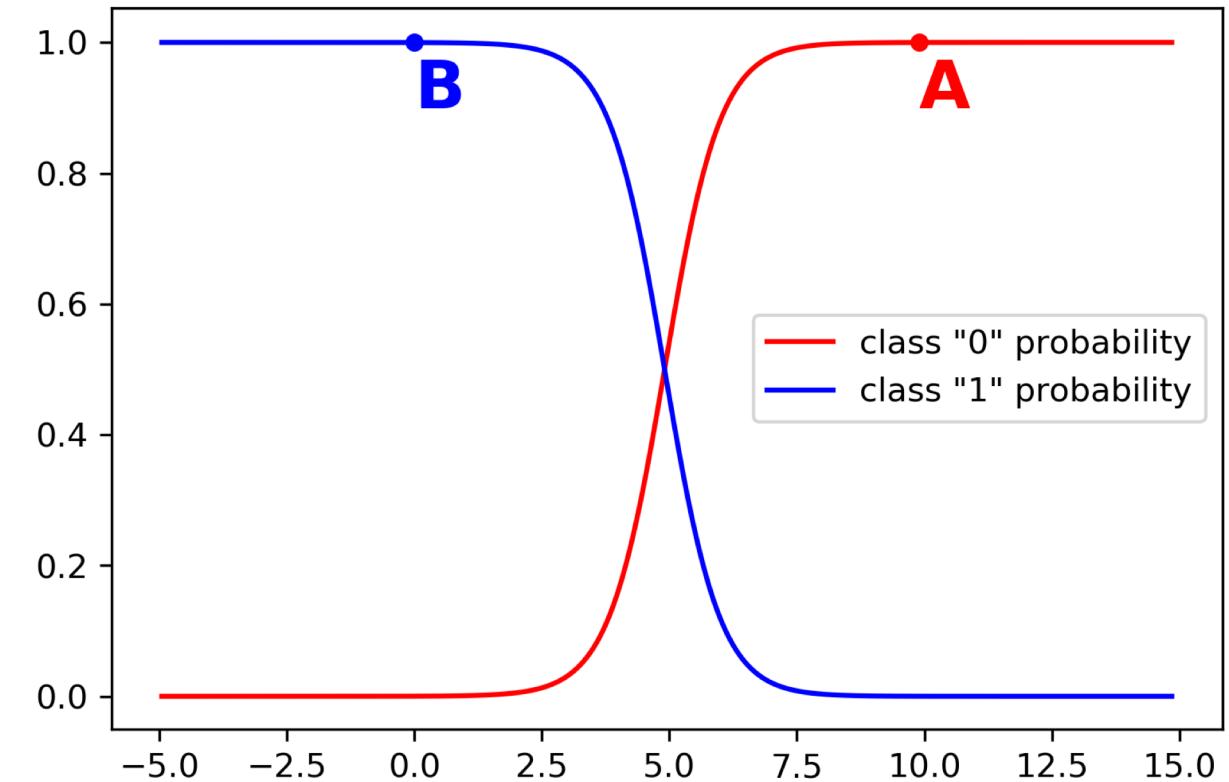
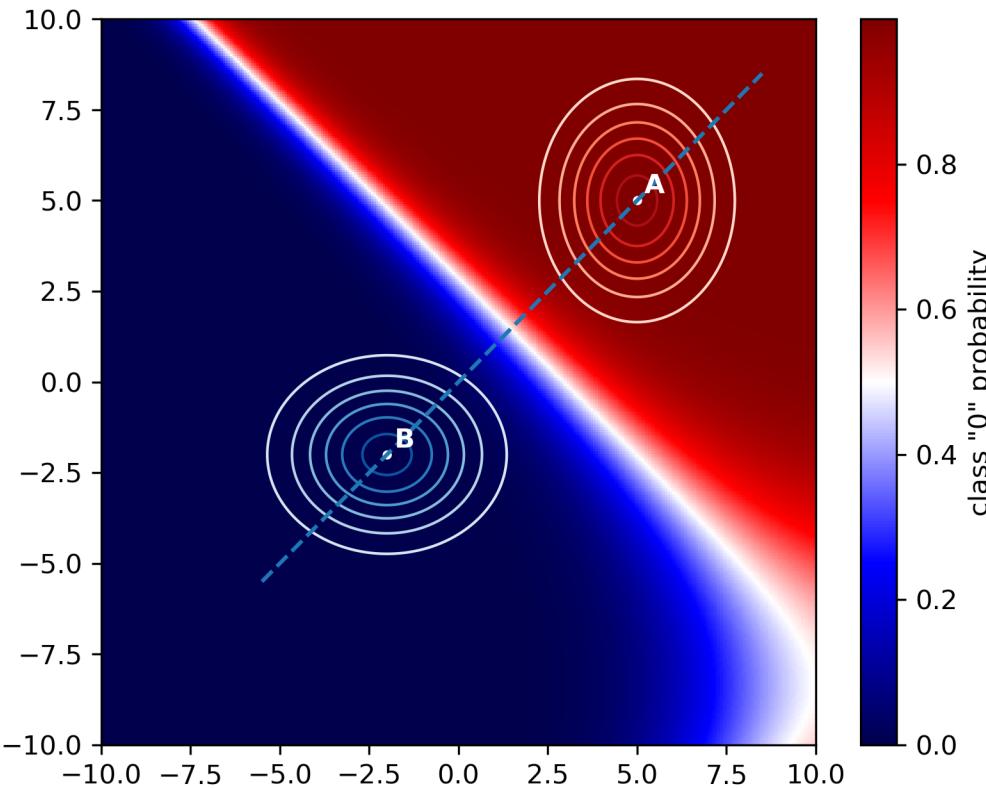


## ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Разделяющая поверхность  
(на примере результатов наивного байесовского классификатора)



# ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации



Может быть, можно аппроксимировать  $P(y = \text{A}|x)$  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 расп-ю Бернулли с параметром  $p$  (зависящий от  $x_i$ );
- (2) Предполагаем, что отношение log-вероятностей классов «1» и «0» соотносятся как линейная функция:

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

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

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

# ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

## Логистическая регрессия («logistic regression»)

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

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

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

Бинарная перекрестная энтропия (отрицательная)  
Negative binary cross-entropy

$$\begin{aligned}\ell(\mathcal{T}, \theta) &= \log(L(\mathcal{T}, \theta)) = \sum_i \log(p_i^{y_i}) + \sum_i \log((1 - p_i)^{(1-y_i)}) = \\ &= \sum_i (y_i * \log p_i + (1 - y_i) * \log(1 - p_i)) = \sum_i \log(1 - p_i) + \sum_i y_i * \log \frac{p_i}{1-p_i} = (*)\end{aligned}$$

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

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

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

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

# ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

## Логистическая регрессия («logistic regression»)

Линейная регрессия:

набор предположений “LINE”

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

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

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

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

Логистическая регрессия:

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

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

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

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

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

# ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

## Логистическая регрессия («logistic regression»)

code demonstration

# ОБУЧЕНИЕ С УЧИТЕЛЕМ: задача классификации

Разделяющая поверхность  
(на примере результатов логистической регрессии)

