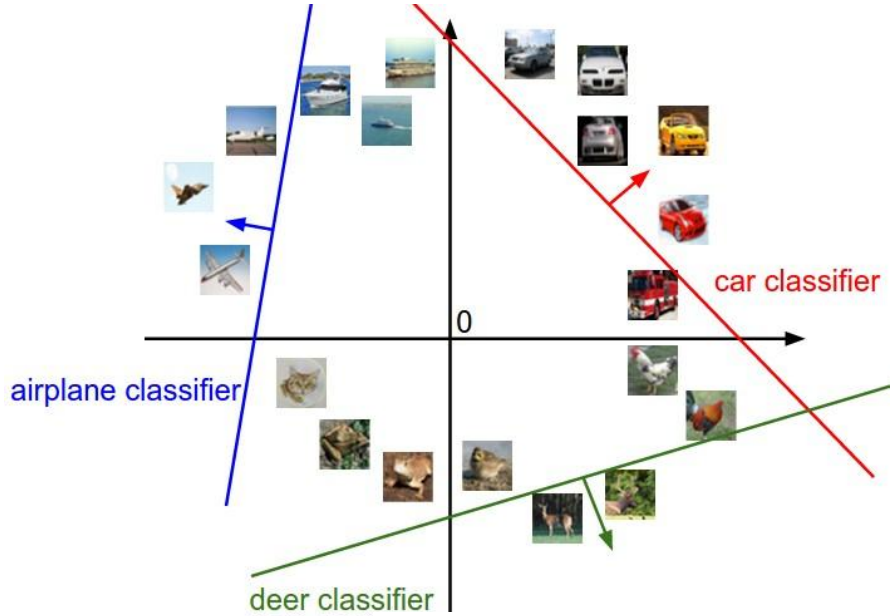
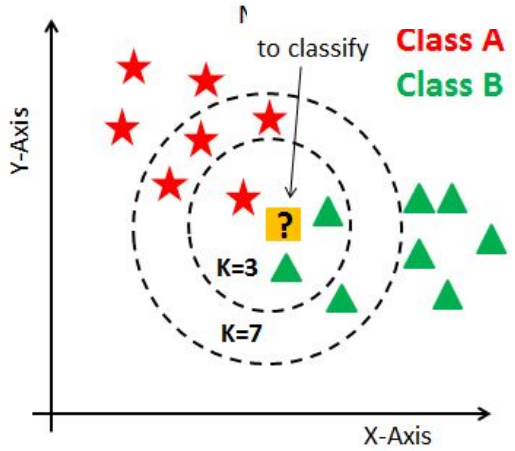
Передовая инженерная аэрокосмическая школа

Искусственный интеллект и машинное обучение

Лекция *02*

Постановка задачи распознавания образов.
Модель классификатора.
Байесовский классификатор.
Построение границы классов.
Разделяющая гиперплоскость.
Метод наименьших квадратов.
Линейная регрессия.
Метод к – ближайших соседей.





https://do.ssau.ru/moodle/course/view.php?id=1459

Постановка задачи распознавания образов

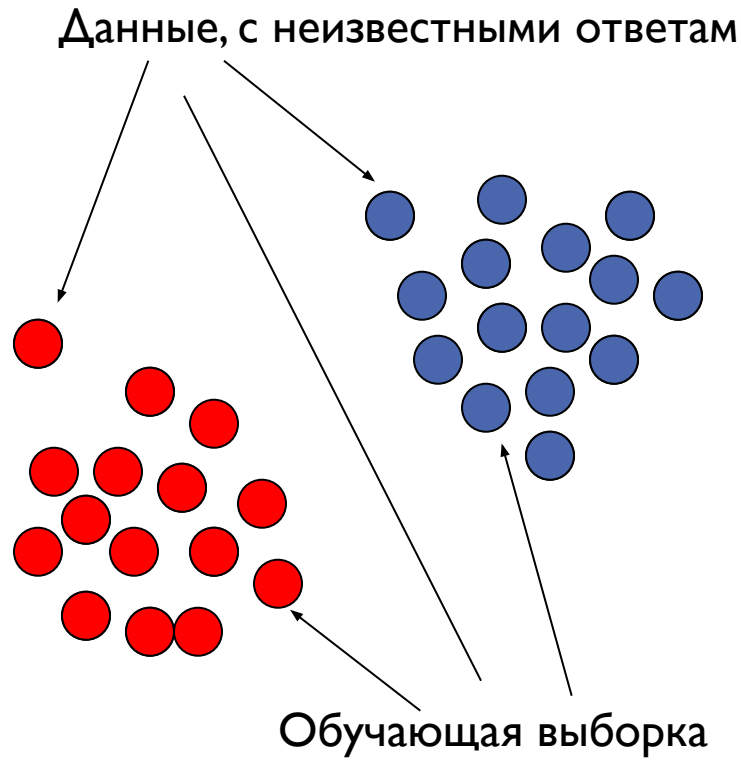
Задача машинного обучения с учителем

- Пусть существуют два множества:
  - Множество *объектов – образов* X
  - Множество *ответов* Y
- Существует целевая функция $y^* : X \to Y$

значения которой известны только на конечном подмножестве объектов $\{x_1, \ldots, x_n\} \subset X$

- Совокупность пар «объект-ответ» $X^N = (x_i, y_i)_{i=1}^N$ – обучающая выборка.

Задача обучения заключается в том, чтобы по выборке X^N построить решающую функцию , $a : X \to Y$которая бы приближала целевую функцию , причём не только на объектах обучающей выборки, но и на всем множестве



Модель алгоритма

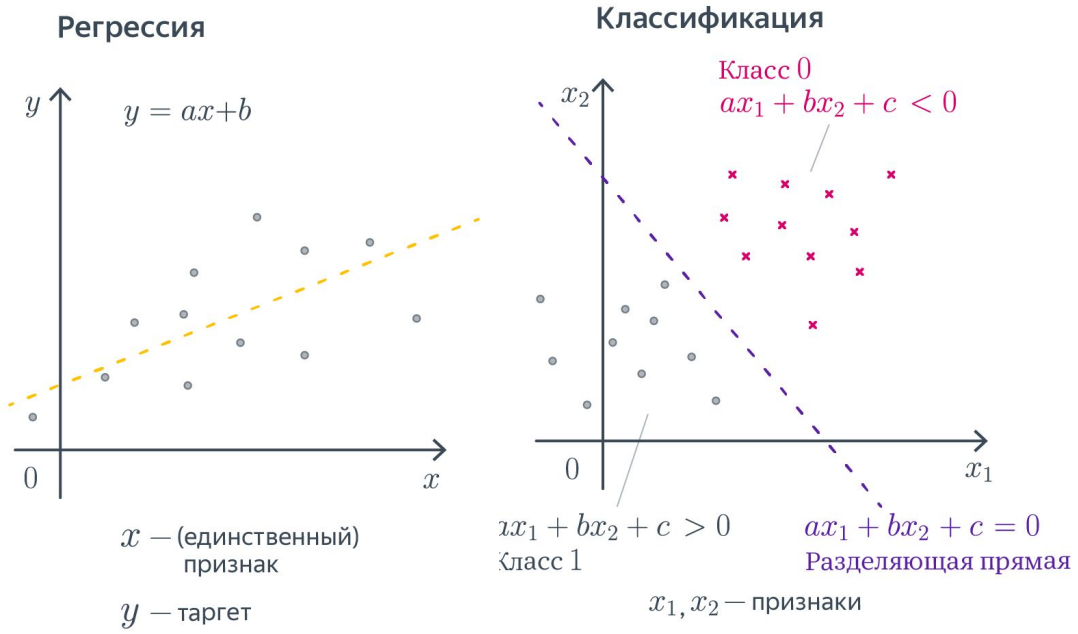
- Требуется построить отображение (гипотезу) $a : X \to Y$
- Пусть A – параметрическое семейство отображений

$$A = \{a(x,\gamma) \mid a : X \times \Gamma \to Y\}$$

  - Г – пространство допустимых значений параметра γ (*пространство поиска*)
- Будем выбирать отображение для решение задачи из A
  - Процесс выбора – ***обучение***
  - Построение отображения $\mu : X^l \to a$

по обучающей выборке – ***метод бучения***

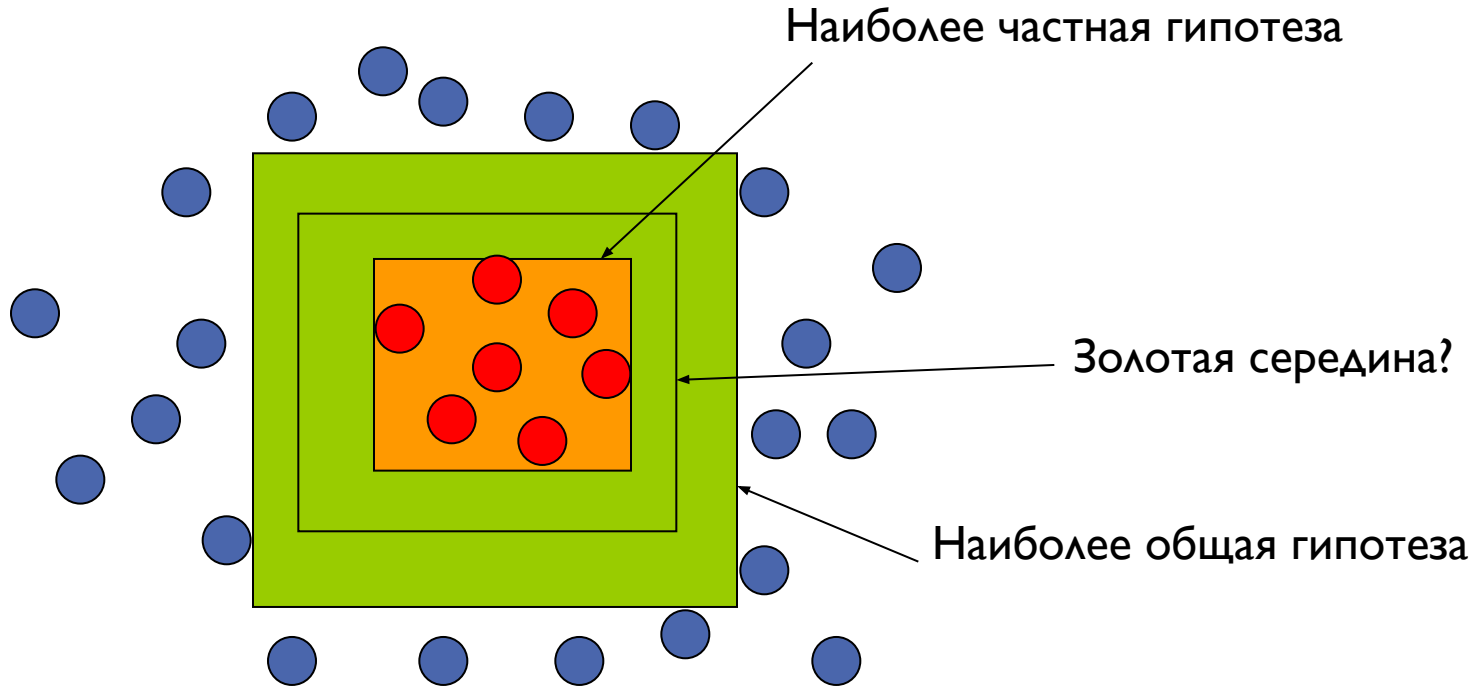
  - Обучение сводится к поиску точки в пространстве поиска Γ



Простейшая модель: $A = \{sign[(\gamma, x) + c]\}_{\gamma,c}$
Пространство поиска – значения вектора γ и смещения c
Гипотеза a – какая-то конкретная прямая

Замечание

- Гипотез, имеющих нулевой эмпирический риск может существовать неограниченное количество:



Вопрос:
Какую модель выбрать?

Эмпирический риск

- $X^l = \{x_1, \ldots, x_l\}$ - обучающая выборка
- Эмпирический риск (ошибка тренировки): $R_{emp}(a, X^l) = \frac{1}{k} \sum_{i=1}^{l} L(a(x_i), y_i)$
- Метод минимизации эмпирического риска*:

$$\mu(X^l) = \arg\min_{a \in A} R_{emp}(a, X^l)$$

$$y' = y^*(x)$$

$$y = a(x)$$

Таким образом задача машинного обучения сводится к задаче оптимизации

$L(a, x)$ – функция потерь = величине ошибки алгоритма a на объекте

$L = Y \times Y \to R$ характеризует отличие правильного ответа от ответа данного построенным отображением

$L(y, y') = [y \neq y']$ - индикатор потери,

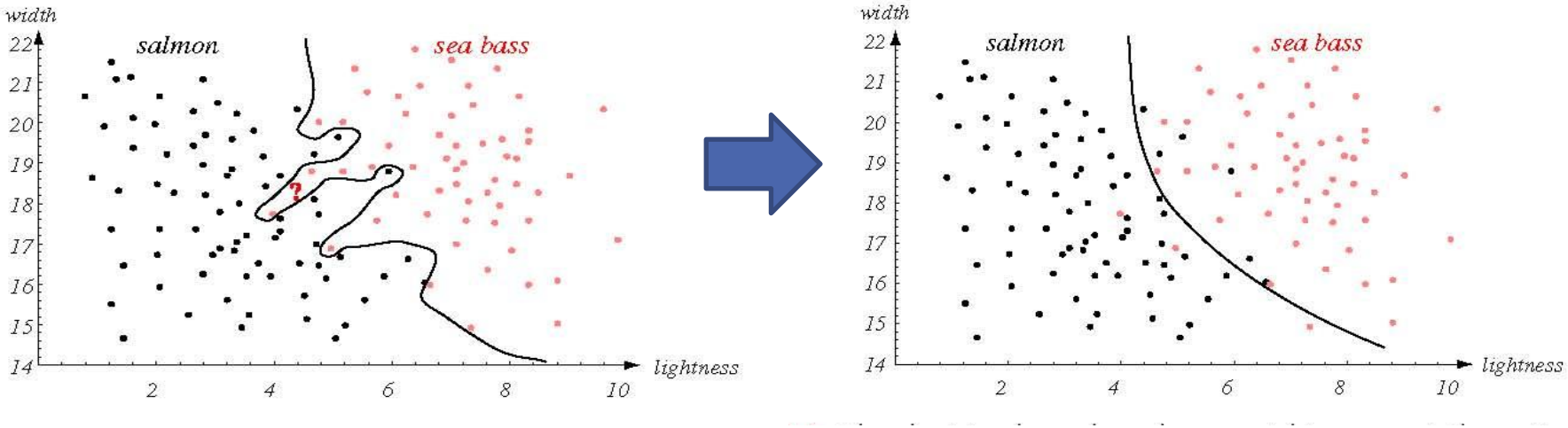
$L(y, y') = |y - y'|$ - отклонение

$L(y, y') = (y - y')^2$ - квадратичное отклонение

$L(y, y') = [|y - y'| > \delta]$ - индикатор существенного отклонения

Обобщающая способность

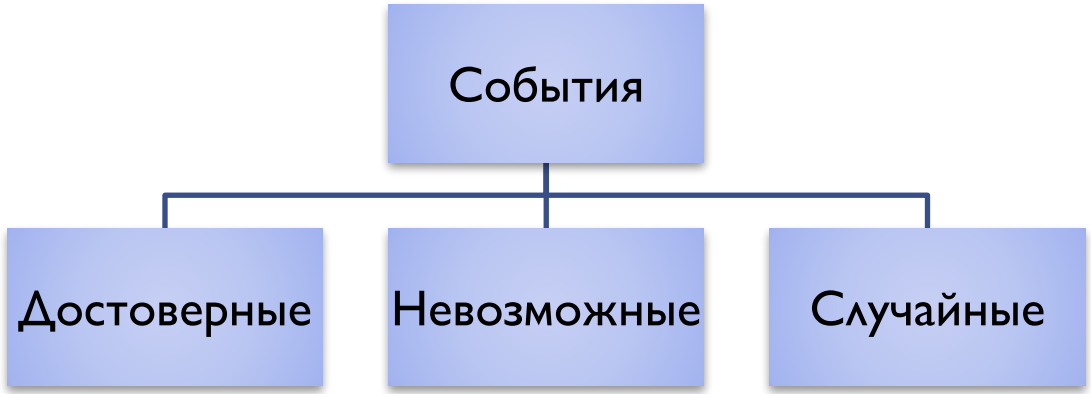
- **Обобщающая способность** (generalization ability, generalization performance).
- Алгоритм обучения обладает *способностью к обобщению*, если вероятность ошибки на тестовой выборке достаточно мала или хотя бы предсказуема, то есть не сильно отличается от ошибки на обучающей выборке.
  - **Проблема обобщения**: малый **эмпирический риск R_{emp}** не означает, что **истинный ожидаемый риск R** будет мал



complex model

simpler model

Основы теории вероятностей: Виды событий

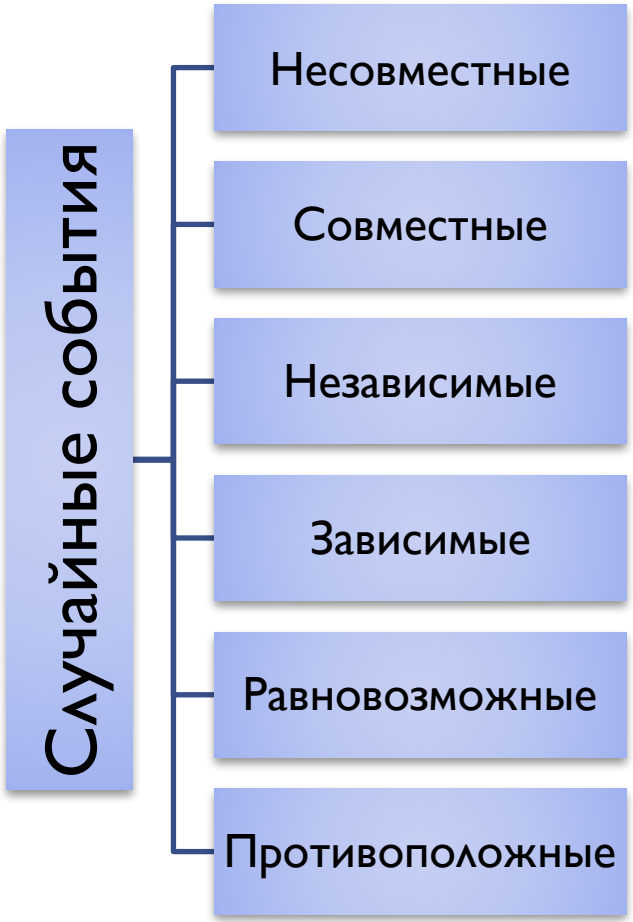


Достоверные события **всегда** происходят при осуществлении данной совокупности условий

Невозможные события **никогда не** происходят при осуществлении данной совокупности условий

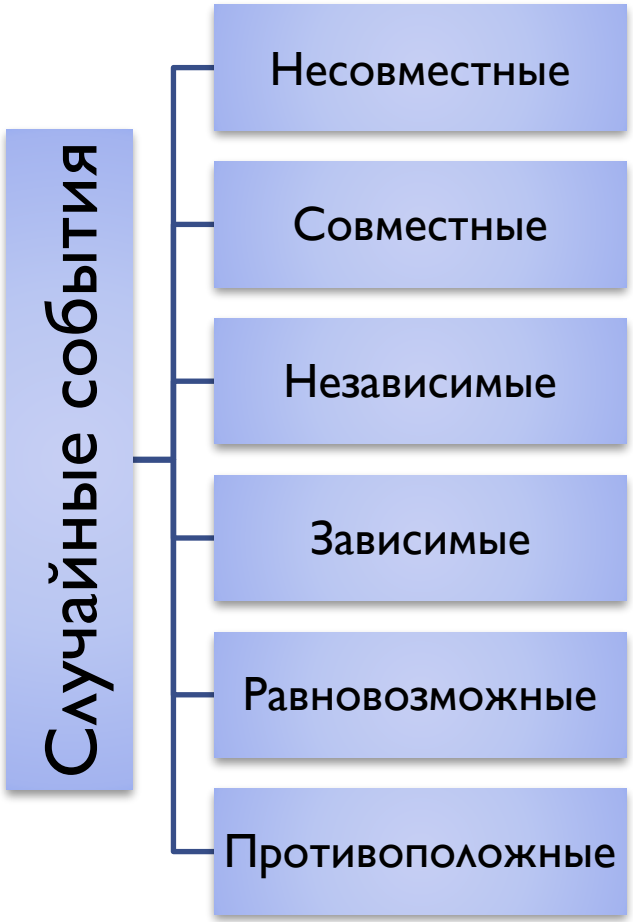
Случайные события **могут произойти** или **не произойти** при осуществлении данной совокупности условий

Основы теории вероятностей: Случайные события



- **Несовместными** называются события, которые не могут одновременно произойти в одном испытании
- Совокупность случайных событий A_1, A_2, A_3, … A_n называется **полной группой** для данного испытания, если в результате испытания обязательно происходит только одно из событий этой совокупности
- Два события (A и $\bar{A}$) называются **противоположными**, если появление одного из них равносильно непоявлению другого

Основы теории вероятностей: Случайные события



- **Совместными** называются события, которые могут одновременно произойти в одном испытании
- События называются **независимыми**, если появление одного из них не изменяет вероятности появления второго.
- События называются **зависимыми** если появление одного из них зависит от появления другого
- **Равновозможными** называются события, если ни у одного из них нет объективного преимущества перед другим

Основы теории вероятностей:
Классическое определение вероятности

Вероятностью события A называют отношение числа благоприятствующих этому событию элементарных событий (m) к общему числу всех равновозможных несовместных элементарных событий (n), образующих полную группу:

$$P(A) = \frac{m}{n}$$

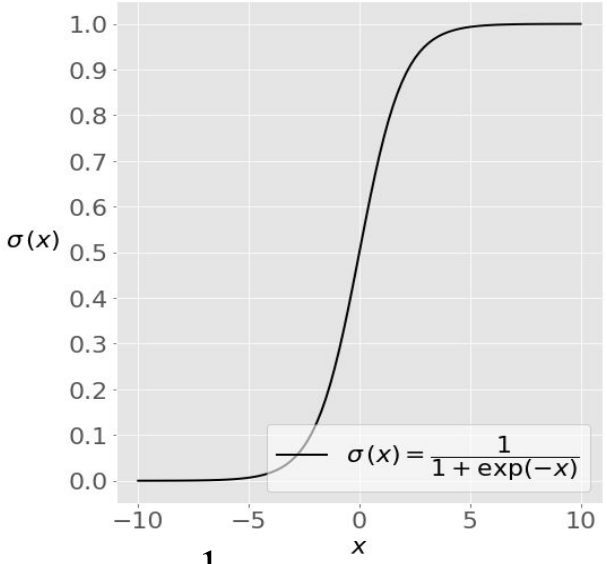
Чтобы рассчитать классическую вероятность необходимо до проведения испытаний теоретически подсчитать:

- общее число всех **равновозможных несовместных** элементарных событий (n)
- число **благоприятствующих** этому событию равновозможных несовместных элементарных событий (m)

Вероятность достоверного **события** $P = 1$
Вероятность невозможного **события** $P = 0$
Вероятность случайного **события** $0 < P < 1$

Логистическая регрессия

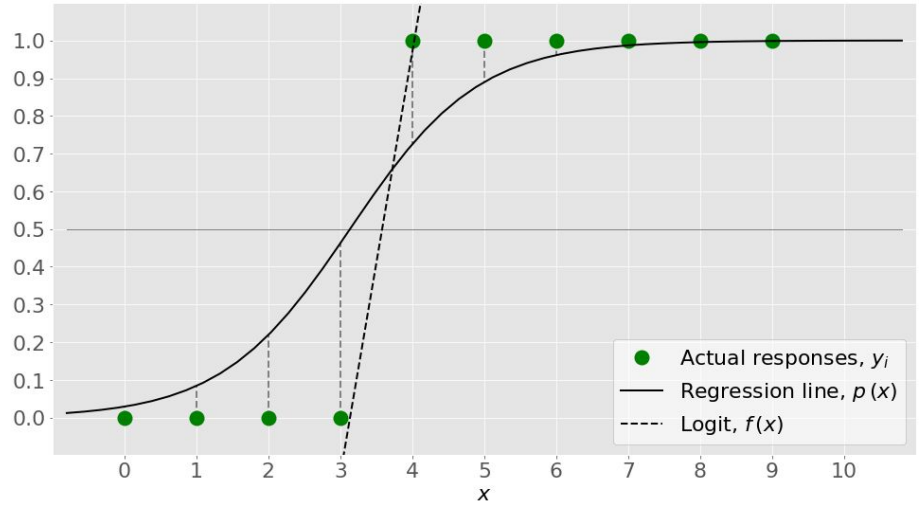
- Статистическая модель, используемая для прогнозирования вероятности возникновения некоторого события путём его сравнения с логистической кривой. Эта регрессия выдаёт ответ в виде вероятности бинарного события (*1* или *0*).
  - Для этого вводится так называемая зависимая переменная, принимающая лишь одно из двух значений — как правило, это числа *0* (событие не произошло) и *1* (событие произошло), и множество независимых переменных (также называемых признаками, предикторами или регрессорами) — вещественных значений на основе которых требуется вычислить вероятность принятия того или иного значения зависимой переменной.
  - Делается предположение о том, что вероятность наступления события $P\{y=1 \mid x\} = f(z)$



$$f(z) = \frac{1}{1+e^{-z}}$$ – логистическая функция

$$z = \sum_n \theta_n x_n$$ – регрессия

- Для оценки коэффициентов регрессии обычно применяется метод оценки максимального правдоподобия



Томас Байес

Thomas Bayes
(c. 1702 – April 17, 1761)

То́мас Ба́йес (в части источников: Бейес, более точная транскрипция: Бейз, англ. Thomas Bayes [beɪz]) — английский математик, пресвитерианский священник, член Лондонского королевского общества (1742).

Математические интересы Байеса относились к теории вероятностей. Он сформулировал и решил одну из основных задач этого раздела математики (теорема Байеса). Работа, посвящённая этой задаче, была опубликована в 1763 году, посмертно. Формула Байеса, дающая возможность оценить вероятность событий эмпирическим путём, играет важную роль в современной математической статистике и теории вероятностей.

Условная вероятность

Определение.

Пусть $P(A)>0$.

Условной вероятностью *P(B/A)* события B при условии, что
событие A наступило, называется число

$$P(B/A)=\frac{P(AB)}{P(A)}$$

Обозначения:
$$P(B/A)=P_A(B)$$

Условная вероятность удовлетворяет всем аксиомам вероятности.

В частности, $0\le P(B/A)\le 1,\ \ P(A/A)=1$

Независимые события

Определение.

- События A и B называются **независимыми**, если

$$P(AB) = P(A)P(B)$$

Определение. Пусть $P(A)>0$ и $P(B)>0$.

- **Событие A не зависит от B**, если $P(A/B) = P(A)$

Следствие.

- Если событие A не зависит от B, то и событие B не зависит от A.

$$P(AB) = P(A/B)P(B) = P(A)P(B)$$

- Доказательство.

$$P(B/A) = \frac{P(AB)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B)$$

На практике из физической независимости событий делают вывод о теоретико-вероятностной независимости.

Полная группа событий

- События $H_1, H_2, \ldots, H_n$ **образуют полную группу**, если они
- 1) попарно несовместны
- 2) в результате эксперимента обязательно какое- либо одно из них наступит $P(H_i H_j) = 0, \ i \neq j$

$$H_1 + H_2 + \ldots + H_n = \Omega$$

H_i - **гипотезы**

Пример.

В стохастическом эксперименте рассмотрим события A и $\overline{A}$

Они образуют полную группу.

Формула полной вероятности

Теорема.

- Если события $H_1, H_2, \ldots, H_n$

образуют **полную группу**,

то для любого события A справедлива формула

$$P(A) = P(H_1)P(A/H_1) + \ldots + P(H_n)P(A/H_n)$$

$$P(A) = \sum_{i=1}^{n} P(H_i)P(A/H_i)$$

Формула Байеса

Теорема.

- Пусть события $H_1, H_2, \ldots, H_n$

образуют **полную группу.**

Пусть событие A наступило $(P(A)>0)$.

Тогда вероятность того,

что при этом была реализована гипотеза (наступило событие) H_k

вычисляется по формуле

$$P(H_k/A) = \frac{P(H_k)P(A/H_k)}{P(A)} = \frac{P(H_k)P(A/H_k)}{\sum_{i=1}^{n} P(H_i)P(A/H_i)}$$

Формула Байеса позволяет переоценить вероятности гипотез после того, как проведено испытание, в результате которого произошло событие A.

Формула Байеса. Частный случай

- Рассмотрим события H и $\overline{H}$

они образуют **полную группу.**

Пусть событие A наступило $(P(A)>0)$.

Тогда вероятность того,

что при этом была реализована гипотеза H

вычисляется по формуле

$$P(H/A)=\frac{P(H)P(A/H)}{P(A)}$$

$$P(H/A)=\frac{P(H)P(A/H)}{P(H)P(A/H)+P(\overline{H})P(A/\overline{H})}$$

Пример:
Какова вероятность увидеть на улице динозавра?

Идя по улице вы видите такую сцену:

(это и есть наблюдение X)

Вычислим вероятность того, что наблюдая такую сцены мы действительно видим динозавра

$$P(y \mid x) = \frac{P(x \mid y) \cdot P(y)}{P(x)}$$

Априорная вероятность встретить динозавра

Правдоподобие – вероятность того, что будь это действительно динозавр наблюдение было бы таким

Априорная вероятность увидеть такую сцену

Пример:
Какова вероятность увидеть на улице динозавра?

Идя по улице вы видите такую сцену:

(это и есть наблюдение X)

Вычислим вероятность того, что наблюдая такую сцены мы действительно видим динозавра

$$P(y \mid x) = \frac{P(x \mid y) \cdot P(y)}{P(x)}$$

Априорная вероятность встретить динозавра

Пусть

$P(x \mid y) = 0.7$ $P(x) = 0.1$ $P(y) = 0.000001$

Правдоподобие – вероятность того, что будь это действительно динозавр наблюдение было бы таким

$$P(y \mid x) = \frac{0.7 \cdot 0.0000001}{0.1} = 0.000007 =$$

Априорная вероятность увидеть такую сцену

Вероятностная формулировка задачи машинного обучения

- Эмпирический риск:

$$R_{Emp}(a, X^l) = P\big(a(x) \neq y \mid X^l\big) = \frac{1}{l}\sum_{i=1}^{l}[a(x_i) \neq y_i]$$

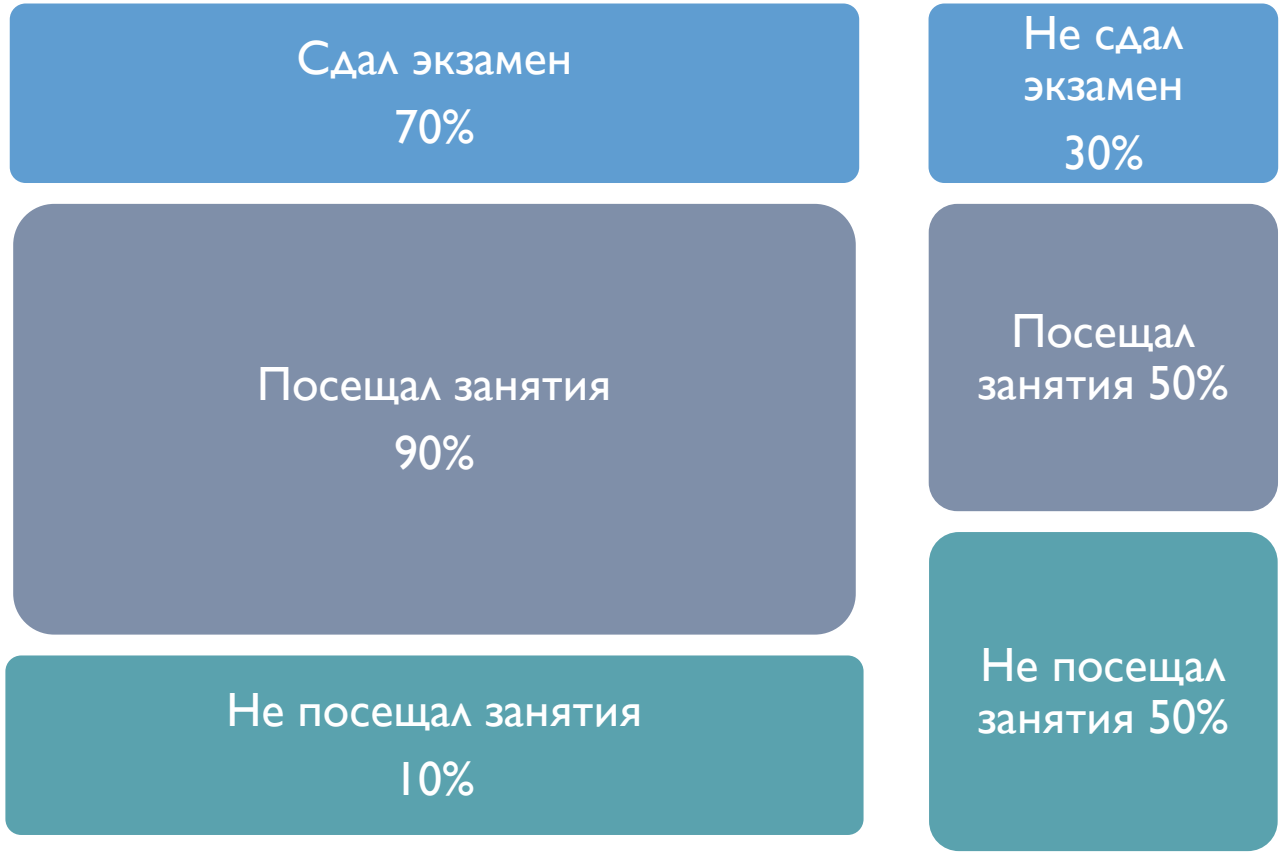
- Общий риск:

$$R(a, X) = P\big(a(x) \neq y \mid X\big) = \int_X P(x)\big[a(x) \neq y\big]dx$$

  - рассчитать невозможно
  - требуется минимизировать

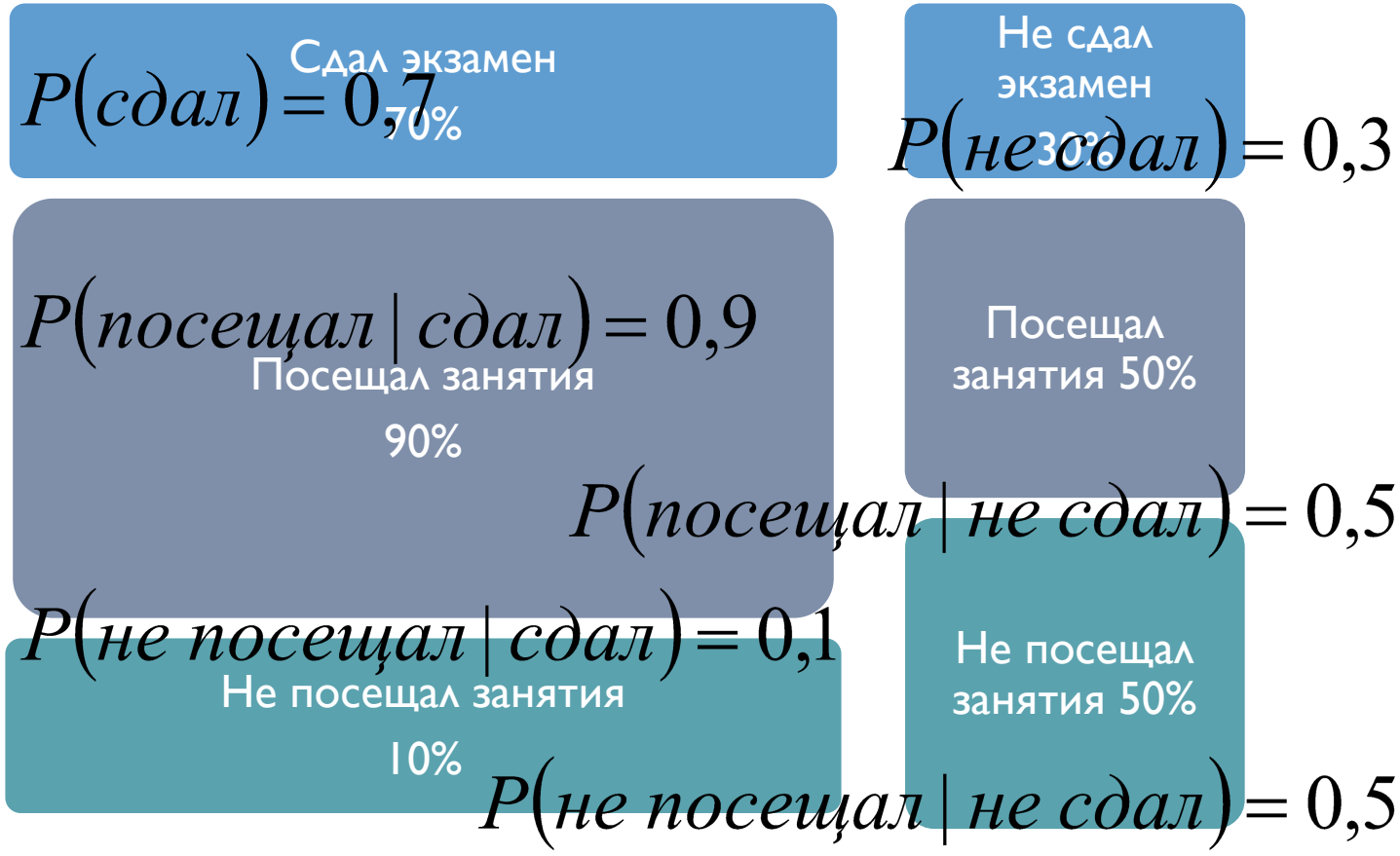
- Модель алгоритма и метод обучения определяются так же

Пример расчёта вероятности



$$P(\text{сдал экзамен} \mid \text{посещал занятия}) = ?$$

Пример расчёта вероятности



$$P(сдал\ экзамен \mid посещал\ занятия) = ?$$

Пример расчёта вероятности

$$P(\text{сдал}) = 0,7 \qquad P(\text{не сдал}) = 0,3$$

$$P(\text{посещал} \mid \text{сдал}) = 0,9$$

$$P(\text{посещал} \mid \text{не сдал}) = 0,5$$

$$P(\text{не посещал} \mid \text{сдал}) = 0,1$$

$$P(\text{не посещал} \mid \text{не сдал}) = 0,5$$

$$P(\text{сдал экзамен} \mid \text{посещал занятия}) =$$

$$= P(\text{сдал})P(\text{посещал} \mid \text{сдал}) / P(\text{посещал}) =$$

$$= 0,7 * 0,9 / 0,78 = 0,81$$

$$P(\text{посещал}) = P(\text{сдал})P(\text{посещал} \mid \text{сдал}) +$$

$$+ P(\text{не сдал})P(\text{посещал} \mid \text{не сдал}) =$$

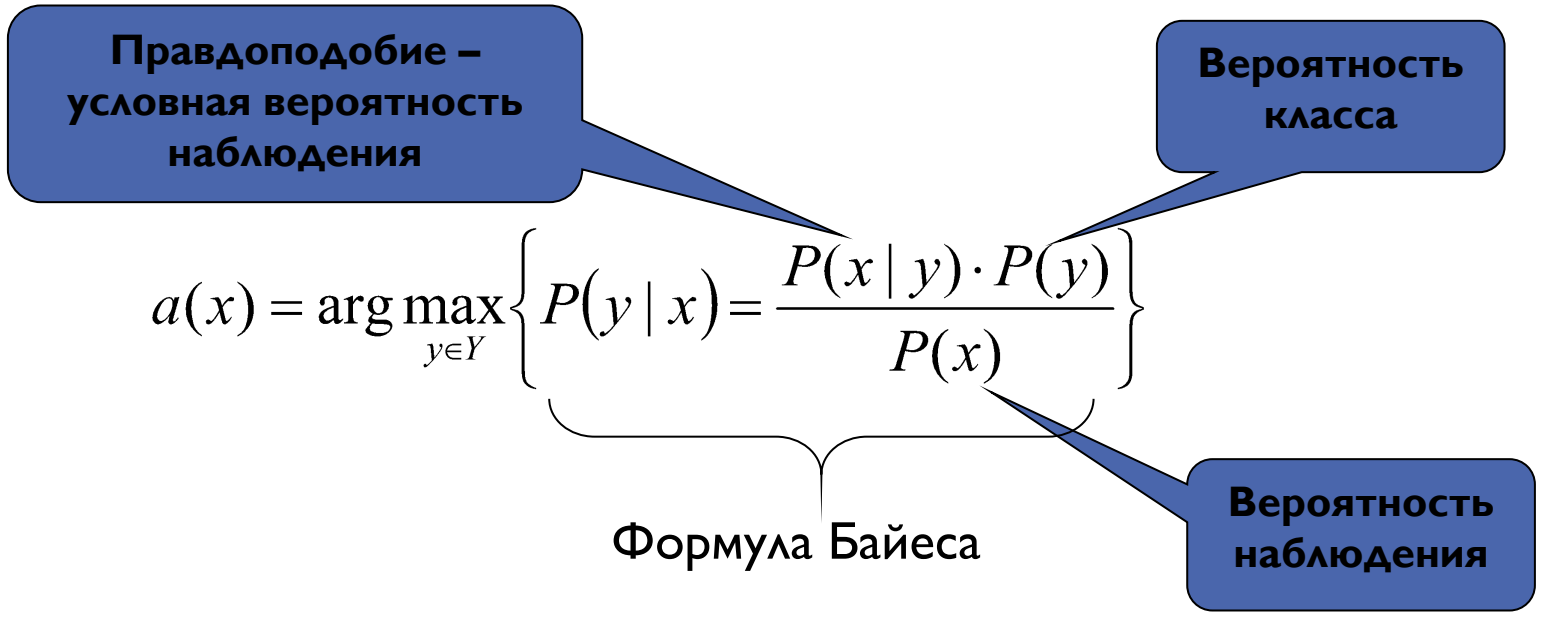
$$= 0,9 * 0,7 + 0,5 * 0,3 = 0,63 + 0,15 = 0,78$$

Домашнее задание 1: Пример расчёта вероятности

- Пусть некий тест на какую-нибудь болезнь имеет вероятность успеха 95%
  - 5% — вероятность как позитивной, так и негативной ошибки.
- Всего болезнь имеется у 1% респондентов.
- Пусть некий человек получил позитивный результат теста
  - тест говорит, что он болен.
- С какой вероятностью он действительно болен?
- Ответ на «Домашнее задание 3.1» разместить на странице курса.

Наивный байесовский классификатор

- **Наи́вный ба́йесовский классифика́тор** — простой вероятностный классификатор, основанный на применении Теоремы Байеса со строгими (наивными) предположениями о независимости.
  - Предположения:
    - Известна функция правдоподобия: $P(x \mid y)$
    - Известны априорные вероятности: $P(y), P(x)$
  - Принцип максимума апостериорной вероятности:



Алгоритм:
Для каждой гипотезы вычислить апостериорную вероятность.
Выбрать гипотезу с максимальной апостериорной вероятностью

Эмпирический риск:

$$R_{emp}(a, X) = P(a(x) \neq y \mid X)$$

Example. Play Tennis

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$\mathbf{X}$=(Sunny, Cool, High, Strong)

Example. Learning.

Outlook	Play=*Yes*	Play=*No*
Sunny	2/9	3/5
Overcast	4/9	0/5
Rain	3/9	2/5

Temp.	Play=*Yes*	Play=*No*
Hot	2/9	2/5
Mild	4/9	2/5
Cool	3/9	1/5

Hum.	Play=*Yes*	Play=*No*
High	3/9	4/5
Normal	6/9	1/5

Wind	Play=*Yes*	Play=*No*
Strong	3/9	3/5
Weak	6/9	2/5

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$P(\text{Play}=Yes) = 9/14$

$P(\text{Play}=No) = 5/14$

Example.Test

x=(Outlook=*Sunny*, Temperature=*Cool*, Humidity=*High*, Wind=*Strong*)

P(Outlook=*Sunny* | Play=*Yes*) = 2/9
P(Temperature=*Cool* | Play=*Yes*) = 3/9
P(Huminity=*High* | Play=*Yes*) = 3/9
P(Wind=*Strong* | Play=*Yes*) = 3/9
P(Play=*Yes*) = 9/14

P(Outlook=*Sunny* | Play=*No*) = 3/5
P(Temperature=*Cool* | Play==*No*) = 1/5
P(Huminity=*High* | Play=*No*) = 4/5
P(Wind=*Strong* | Play=*No*) = 3/5
P(Play=*No*) = 5/14

$$P(Yes \mid \mathbf{x}) \approx [P(Sunny \mid Yes)P(Cool \mid Yes)P(High \mid Yes)P(Strong \mid Yes)]P(Play=Yes)$$

$$P(No \mid \mathbf{x}) \approx [P(Sunny \mid No)\ P(Cool \mid No)P(High \mid No)P(Strong \mid No)]P(Play=No)$$

$$P(Yes \mid \mathbf{x}) \gtrless P(No \mid \mathbf{x})$$

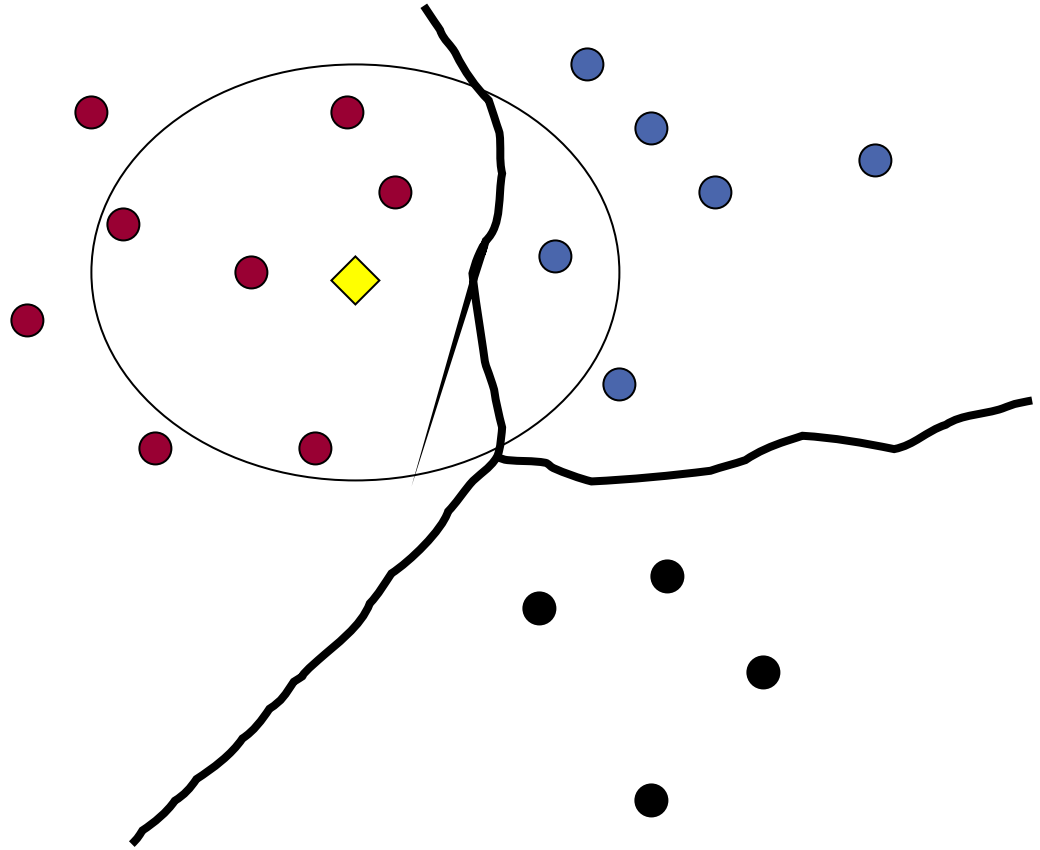
Решение

Домашнее задание 3.2 – для указанного варианта определить вероятность благоприятного и неблагоприятного исхода, сделать вывод!

Особенности наивного байесовского классификатора

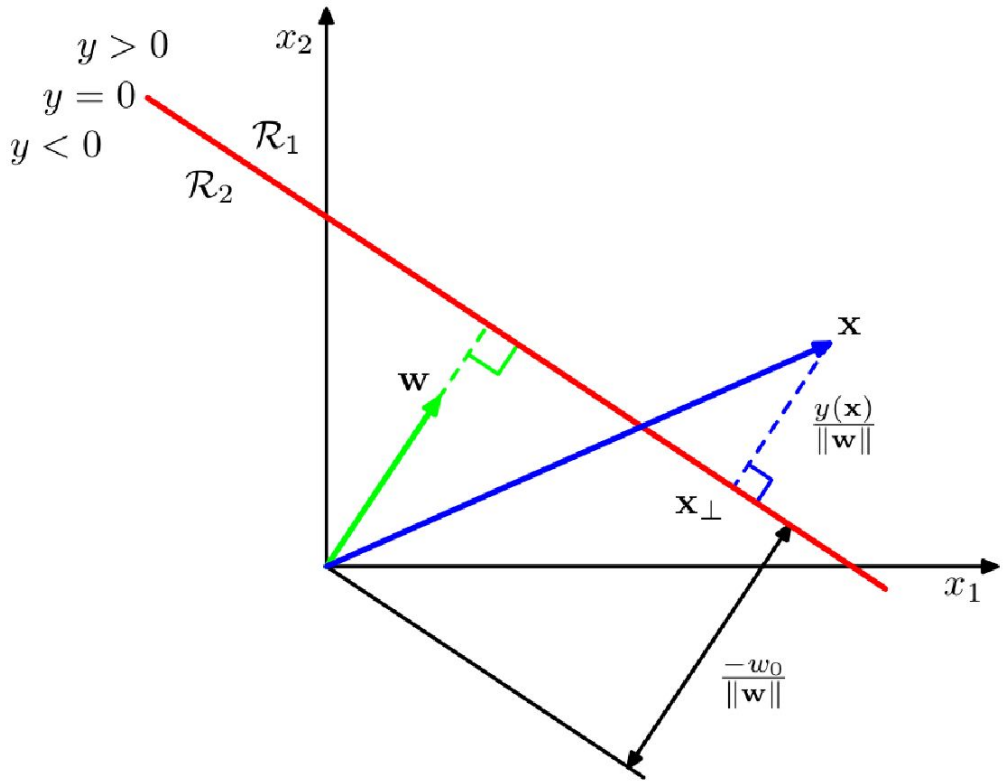
- Нужно знать функцию правдоподобия и априорные вероятности
- *Отсутствуют априорные причины верить, что одна из гипотез более вероятна чем другая (наивность)*
- *Отвечает на вопрос – Какова наиболее вероятная гипотеза при имеющихся данных?*
- *Надо ответить на вопрос – Какова наиболее вероятная классификации нового примера при имеющихся данных?*

Построение границы классов



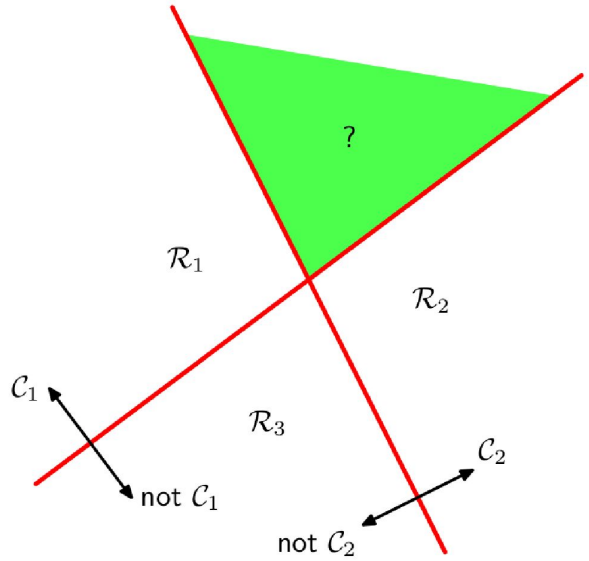
Разбиение пространства, как задача классификации

- Задача классификации: определить вектор x в один из K классов Y .
- В итоге у нас так или иначе всё пространство разобьётся на эти классы.
- Т.е. на самом деле мы ищем разделяющую поверхность (decision surface, decision boundary).
- Рассмотрим линейную дискриминантную функцию: $y(x) = w^T x + w_0$

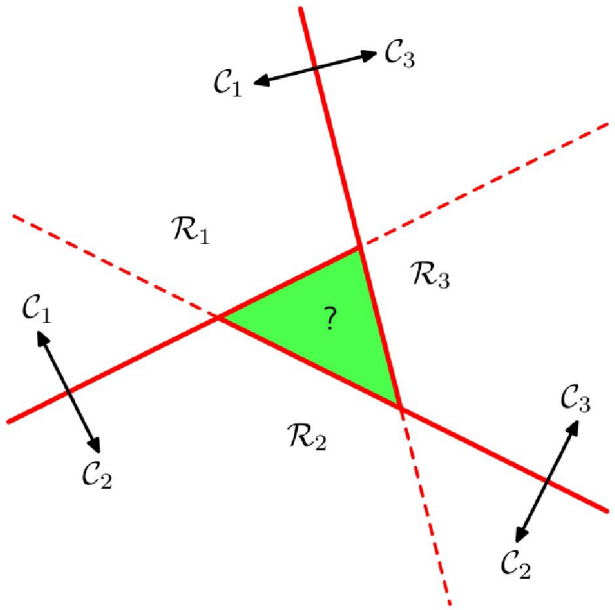


Разделение на несколько классов

- Можно рассмотреть поверхности вида «один против всех»
- Можно рассмотреть поверхности вида «каждый против каждого»
- Можно рассмотреть единый дискриминант из k линейных функций вида



$$y_k(x) = w_k^T x + w_{k0}$$



- Классифицируем в Y_k если соответствующий y_k – максимален

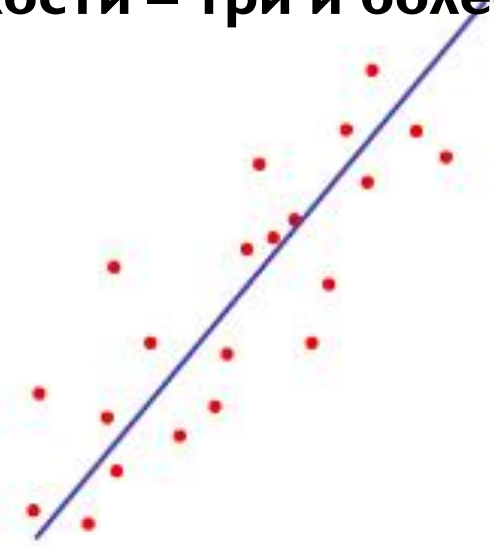
Задача линейной регрессии

- **Нужно найти функцию, которая отображает зависимость одних переменных или данных от других**.
- Зависимые данные называются **зависимыми переменными**, **выходами** или **ответами**.
- Независимые данные называются **независимыми переменными, входами или предсказателями**.
- Обычно в регрессии присутствует одна непрерывная и неограниченная зависимая переменная.
- Входные переменные могут быть неограниченными, дискретными или категорическими данными

Задача линейной регрессии

Через две точки на плоскости можно провести прямую и только одну

А если точек на плоскости – три и более?



Метод наименьших квадратов (МНК) состоит в том, чтобы найти такие коэффициенты регрессии, при которых достигается минимум следующего функционала качества на заданной обучающей выборки

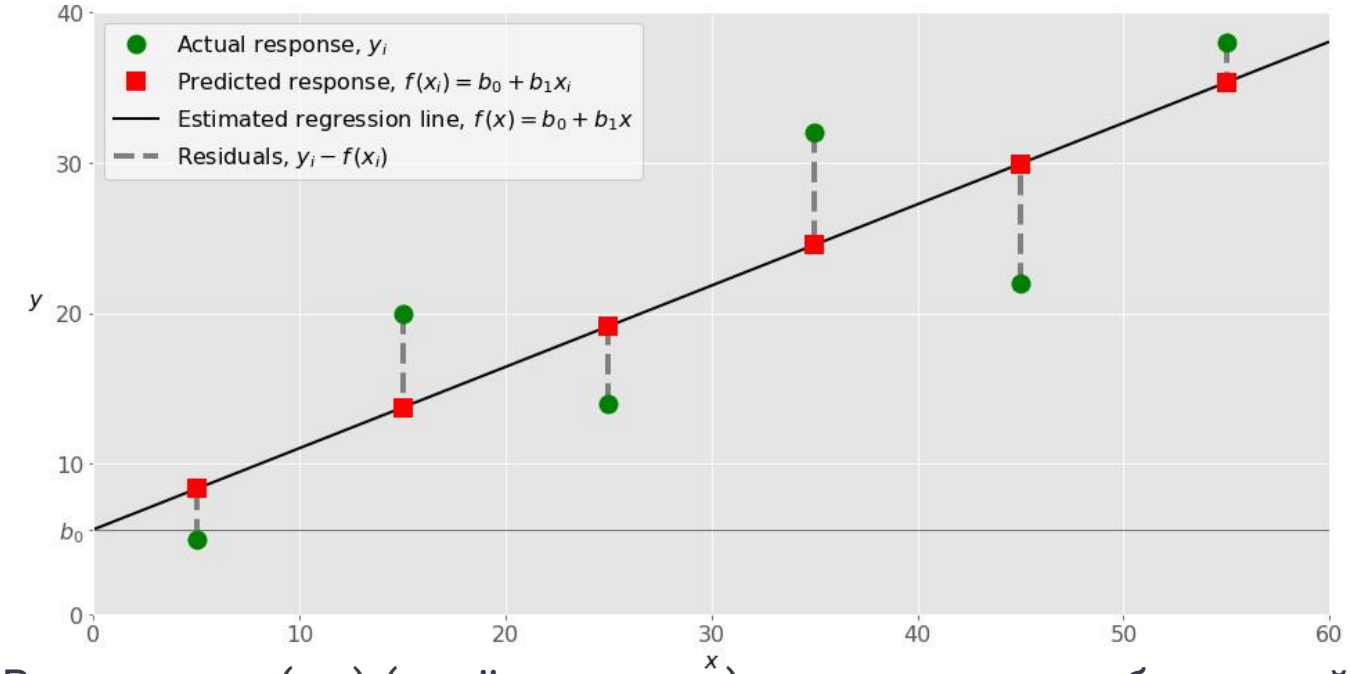
$$y_i' = y_k'(x_i) = w_k^T x_i + w_k^0$$

$$\sum_i (y_i - y_i')^2 = \sum_i (y_i - w_k^T x_i + w_k^0)^2 \xrightarrow{k} \min$$

Scikit-learn

- Библиотека Scikit-learn — самый распространённый выбор для решения задач классического машинного обучения.
- Scikit-learn специализируется на алгоритмах машинного обучения для решения задач
  - обучения с учителем:
    - ***классификации*** *(предсказание признака, множество допустимых значений которого ограничено)*
    - ***регрессии*** *(предсказание признака с вещественными значениями)*
  - обучения без учителя:
    - ***кластеризации*** *(разбиение данных по классам, которые модель определит сама)*,
    - ***понижения размерности*** *(представление данных в пространстве меньшей размерности с минимальными потерями полезной информации)*
    - ***детектирования аномалий***.

Scikit-learn

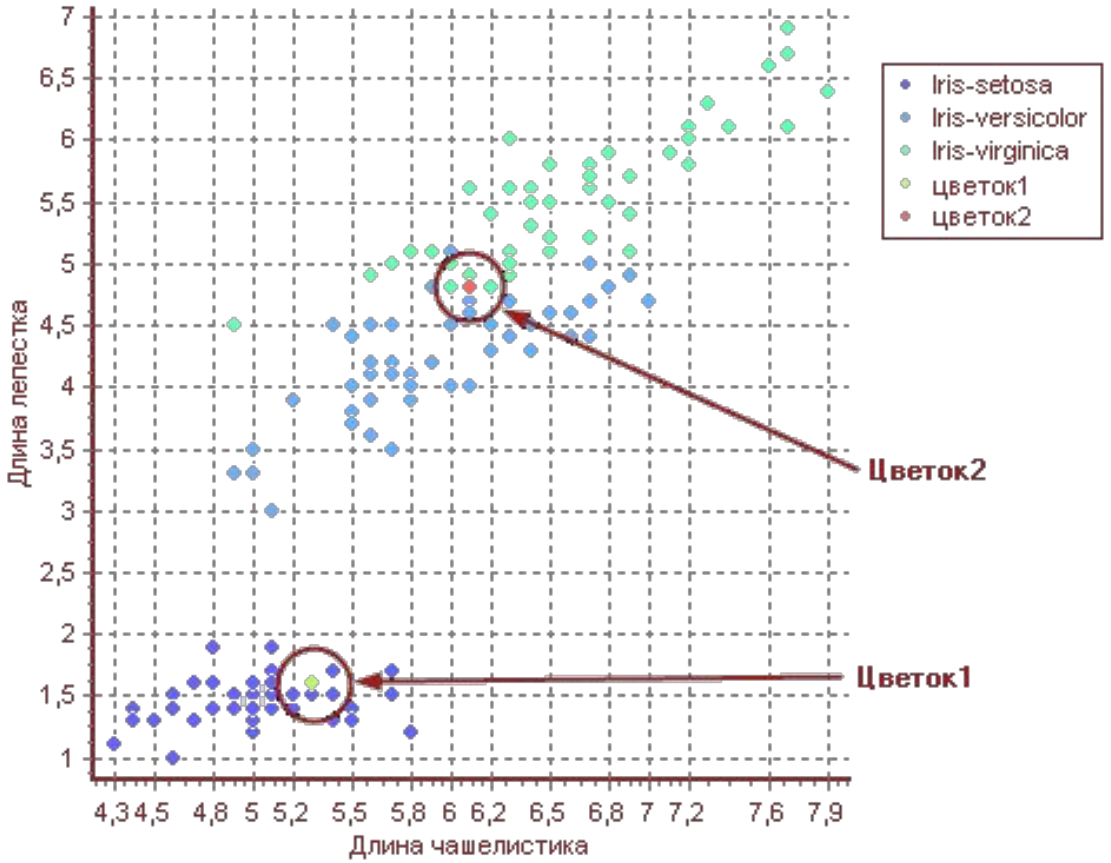


- Вход-выход (x-y) (зелёные круги) – результаты наблюдений.
- Оценочная функция регрессии (чёрная линия) выражается уравнением $f(x) = b_0 + b_1 x$.
- Предсказанные ответы (красные квадраты) – точки линии регрессии, соответствующие входным значениям.
- Остатки (вертикальные пунктирные серые линии) – при реализации линейной регрессии минимизируется сумма квадратов расстояний.

Пример: Ирисы Фишера

150 цветков трех классов:

Два параметра: длина чашелистика и длина лепестка.



Два новых цветка со следующими значениями длины чашелистика и лепестка: 5,3 и 1,6 (*цветок 1*), 6,1 и 4,8 (***цветок 2***).

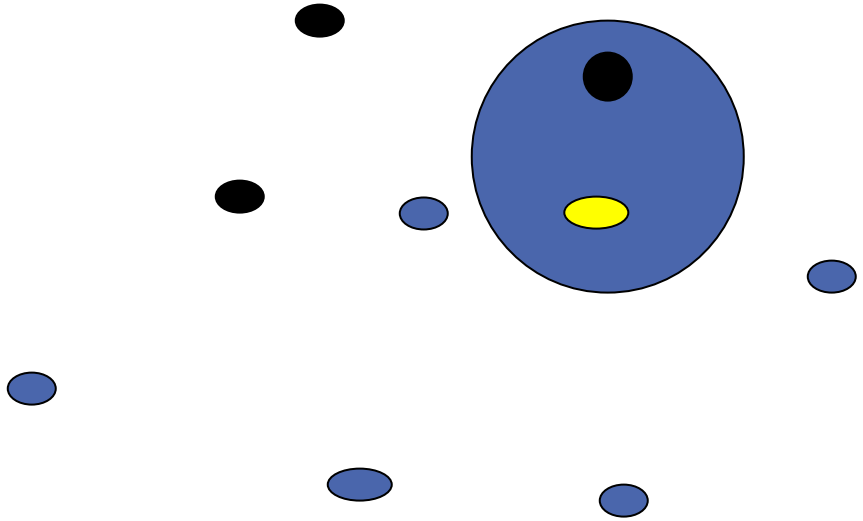
Задание 1. Построить регрессию

Метод «k-ближайших соседей». Классификатор

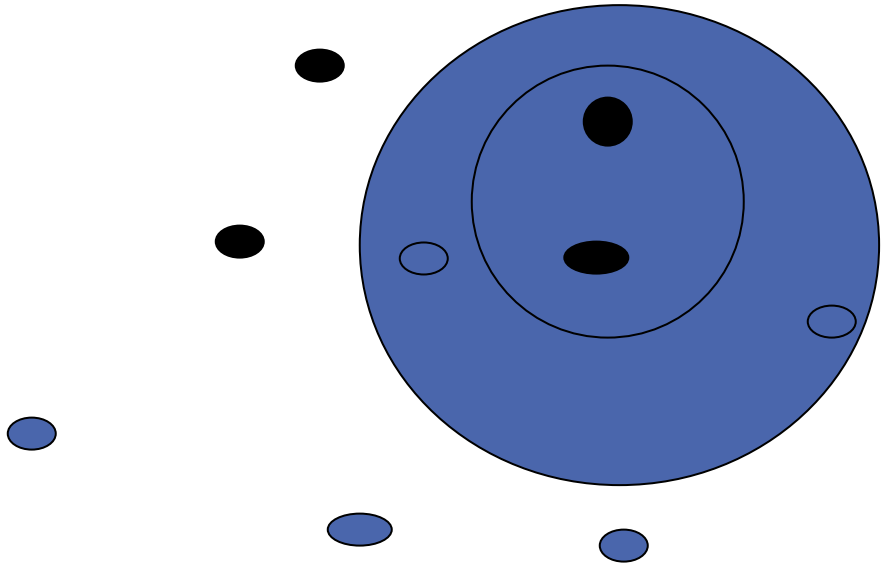
K-nearest neighbor – kNN

- Метод решения задачи классификации, который относит объекты к классу, которому принадлежит большинство из *k* его ближайших соседей в многомерном пространстве признаков.
- Число *k* – это количество соседних объектов в пространстве признаков, которое сравнивается с классифицируемым объектом.
- Использование только одного ближайшего соседа (1NN) ведёт к ошибкам из-за:
  - нетипичных примеров
  - ошибок в ручной привязке единственного обучающего примера.
- Более устойчивой альтернативой является k наиболее похожих примеров и определение большинства
- Величина k типично нечётная: 3, 5

1-Nearest Neighbor



3-Nearest Neighbor



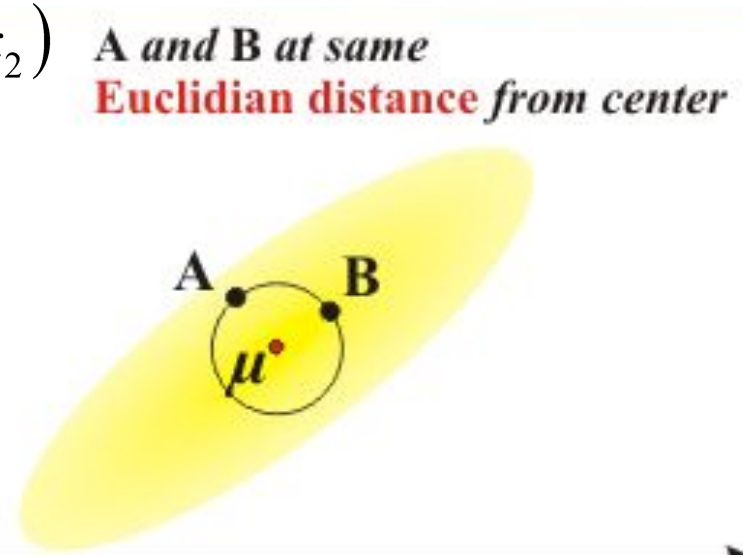
Нормализация и вычисление расстояния

Евклидово расстояние $\rho(x_1, x_2) = (x_1 - x_2)^T (x_1 - x_2)$

Минимаксная нормализация: $x^* = \frac{x - x_{min}}{x_{max} - x_{min}}$

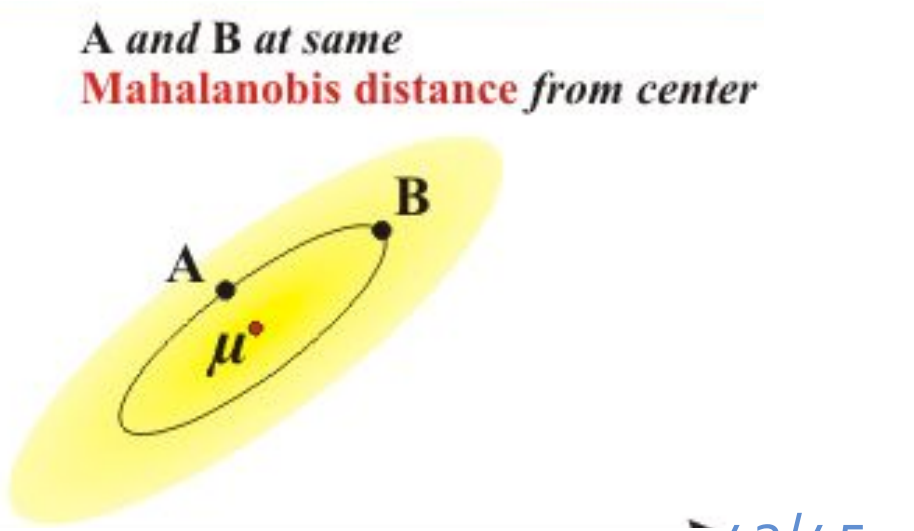
Нормализация с помощью стандартного отклонения: $x^* = \frac{x - x_{mean}}{\sigma_x}$

где σ_x - стандартное отклонение.



Расстояние Махаланобиса

$$\rho(x_1, x_2) = (x_1 - x_2)^T B^{-1} (x_1 - x_2)$$



Предложено индийским статистиком Махаланобисом в 1936 году. С помощью расстояния Махаланобиса можно определять сходство неизвестной и известной выборки. Оно отличается от расстояния Евклида тем, что учитывает корреляции между переменными и инвариантно к масштабу.

Ирисы Фишера: Простое голосование

Цветок 1. Зададим k=3.

Ближайшие соседи: A (5,3; 1,5), B(5,2; 1,5) и C(5,2; 1,5).

$$d(\text{цветок } 1, A) = \sqrt{(5{,}3 - 5{,}3)^2 + (1{,}6 - 1{,}5)^2} = 0{,}1$$

$$d(\text{цветок } 1, B) = \sqrt{(5{,}3 - 5{,}2)^2 + (1{,}6 - 1{,}5)^2} = 0{,}14$$

$$d(\text{цветок } 1, \text{C}) = \sqrt{(5{,}3 - 5{,}2)^2 + (1{,}6 - 1{,}5)^2} = 0{,}14$$

Объект	Чашелистик	Лепесток	Расстояние	Класс
Цветок 1	5,3	1,6	-	-
A	5,3	1,5	0,1	Iris Setosa
B	5,2	1,5	0,14	Iris Setosa
C	5,2	1,5	0,14	Iris Setosa

Класс цветка 1: *Iris Setosa*

Ирисы Фишера: Простое голосование

Цветок 2. Зададим k=3 и предположим, что длина лепестка вдвое важнее длины чашелистика.

Ближайшие соседи: $A(6{,}1;\ 4{,}7)$, $B(6;\ 4{,}8)$, $C(6{,}2\ 4{,}8)$

$$d(\text{цветок } 1, A) = \sqrt{(6{,}1 - 6{,}1)^2 + 2(4{,}8 - 4{,}7)^2} = 0{,}14$$

$$d(\text{цветок } 1, B) = \sqrt{(6{,}1 - 6)^2 + 2(4{,}8 - 4{,}8)^2} = 0{,}1$$

$$d(\text{цветок } 1, C) = \sqrt{(6{,}1 - 6{,}2)^2 + 2(4{,}8 - 4{,}8)^2} = 0{,}1$$

Объект	Чашелистик	Лепесток	Расстояние	Класс
Цветок 2	6,1	4,8	-	-
A	6,1	4,7	0,14	Iris Versicolour
B	6	4,8	0,1	Iris Virginica
C	6,2	4,8	0,1	Iris Virginica

Класс *цветка 2*: *Iris Virginica*

Достоинства kNN

Достоинства

- ✔ Программная реализация относительно проста.
- ✔ Возможность модификации алгоритма.
- ✔ Алгоритм устойчив к аномальным выбросам.
- ✔ Возможность интерпретации результатов работы алгоритма.

Недостатки

- Набор данных, используемый для алгоритма, должен быть репрезентативным.
- Необходимость хранить обучающую выборку целиком.
- В простейших случаях метрические алгоритмы имеют крайне бедный набор параметров, что исключает возможность настройки алгоритма по данным.
- Затраты в производительности велики, поскольку нам необходимо вычислить расстояния между каждым экземпляром и всеми пробными экземплярами.