**Классификация** – цель: спрогнозировать метку класса, которая представляет собой выбор из заранее определенного списка возможных вариантов.

(ирис классифицировали на 3 сорта)

**Классификация:**

1. бинарная – частный случай разделения на 2 класса

(попытка ответить на вопроса в формате «да/нет»)

1. мультиклассовая – в классификации участвует 2 и более участников

(ирис на 3 класса)

**Регрессия –** цель: спрогнозировать непрерывное число (число с плавающей точкой, вещественное число)

(прогнозирование годового дохода чела в зависимости от знаний, образования, места жительства)

Прогнозируемое значение – **сумма** и мб любым значением в заданном диапазоне.

**Отличить классификацию от регрессии?**

Спросить, заложена ли в полученном ответе определенная преемственность. Если полученные результаты непрерывно связаны друг с другом – регрессия.

* Разница между доходами 3999 и 4001 – несущественна – регрессия
* Сайт, который написан на разных языках (между ними нету промежуточного) – классификация

**Способность обобщать** результат на тестовые данные – способность модели выдавать точные прогнозы на ранее не встречавшихся данных.

**Переобучение** – построение модели, которая слишком сложна для имеющегося у нас набора информации.

**Переобучение** происходит, когда модель слишком точно подстраивается под особенности обучающего набора и мы получаем модель, которая хорошо работает на обучающем наборе, но не умеет обобщать результат на новые данные.

**Недообучение** – выбор слишком простой модели.

Используем простые, низко размерные наборы данных, т.к. их легко визуализировать. Данные, которые имеют более 2 признаков графически представить трудно.

Вывод, полученный для **низкоразмерного** (небольшим числом признаков) набора может не подтвердится для **высокоразмерного** (большим количеством признаков) набора данных.

**С учителем.**

**Задание:**

\* построить модель на обучающих данных => получить точные прогнозы для новых, еще не встречающихся, которые имеют те же самые характеристики, что и обучающих набор

\* построить модель, которая будет обладать максимально обобщающей стоимостью