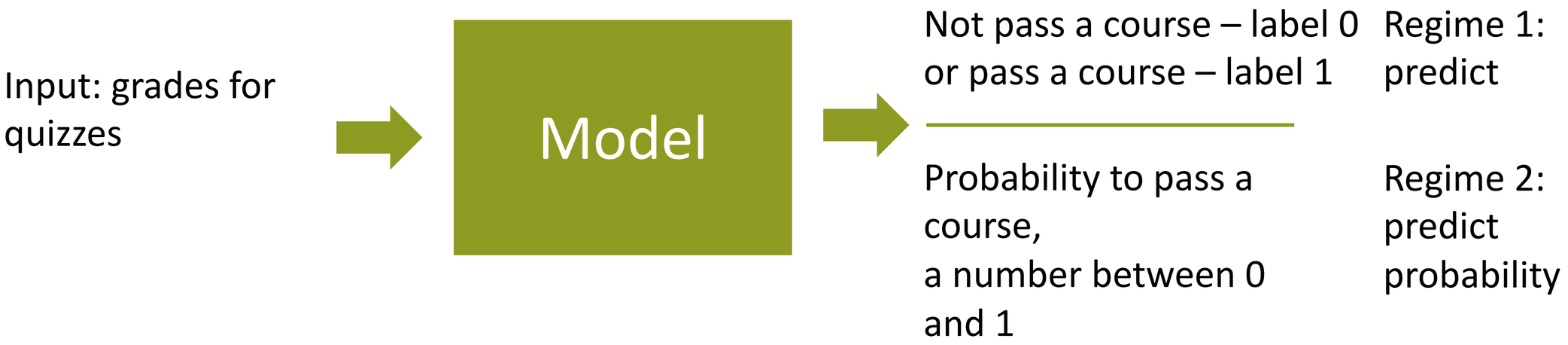
Classification

Profs E. Burnaev
A. Zaytsev
Skoltech

Classification model: can work in two regimes

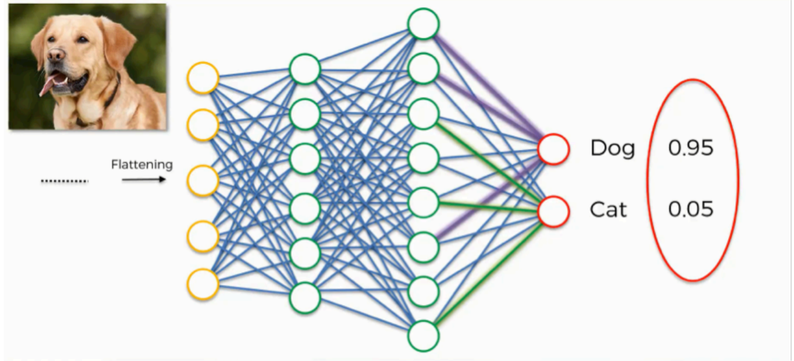
Predict whether a student will pass a course on ML



A Zoo of Classification Methods and how to estimate their accuracy

- ✓ K-nearest neighbors
- ✓ Logistic regression
- ✓ Decision Trees
- ✓ Forest (ensemble) of Decision Trees, etc.

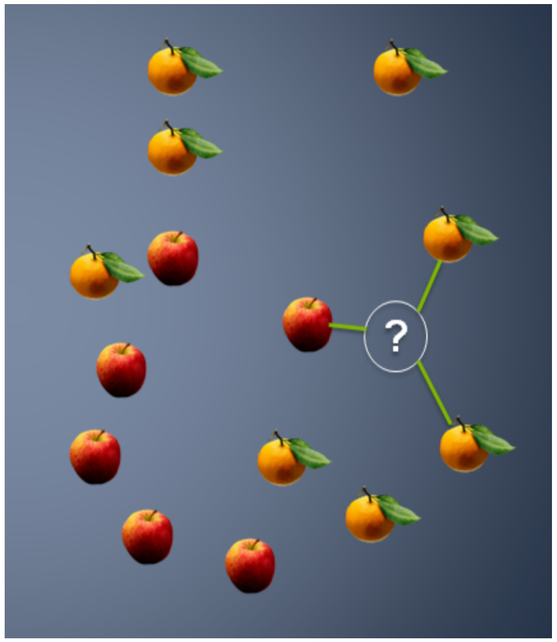
- ➢ Accuracy
- ➢ Errors of the 1st and 2nd kind
- ➢ Conjugacy matrix
- ➢ Precision and recall
- ➢ ROC curve
- ➢ PR curve



k-nearest neighbors

Predict a class of an object by voting between k nearest neighbors

Which prediction we will get for an object (?) for k = 3



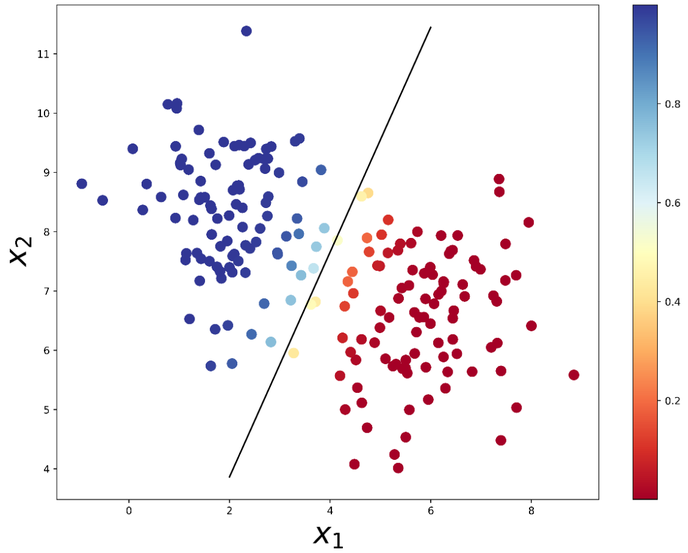
Logistic regression

We would like a model that will predict a number between 0 and 1 – a probability to belong to a positive class.

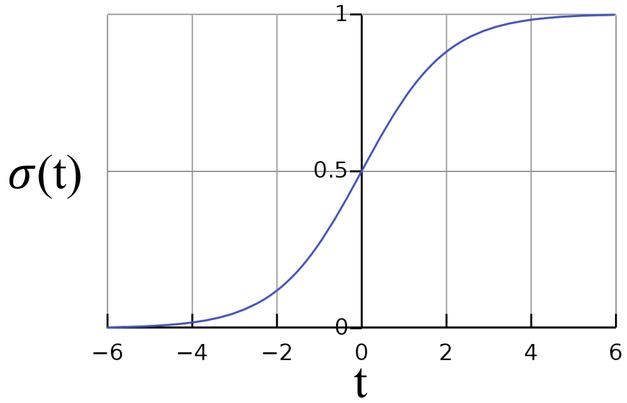
Logistic regression transforms outputs of a linear model so that we predict a probability

$$\hat{y} = \sigma(w_1 x_1 + w_2 x_2 + b)$$

$\hat{y}$ – the probability of default of a borrower
x_1 – age
x_2 – the amount of delayed payments
….



Blue points – first class,
red points – second class

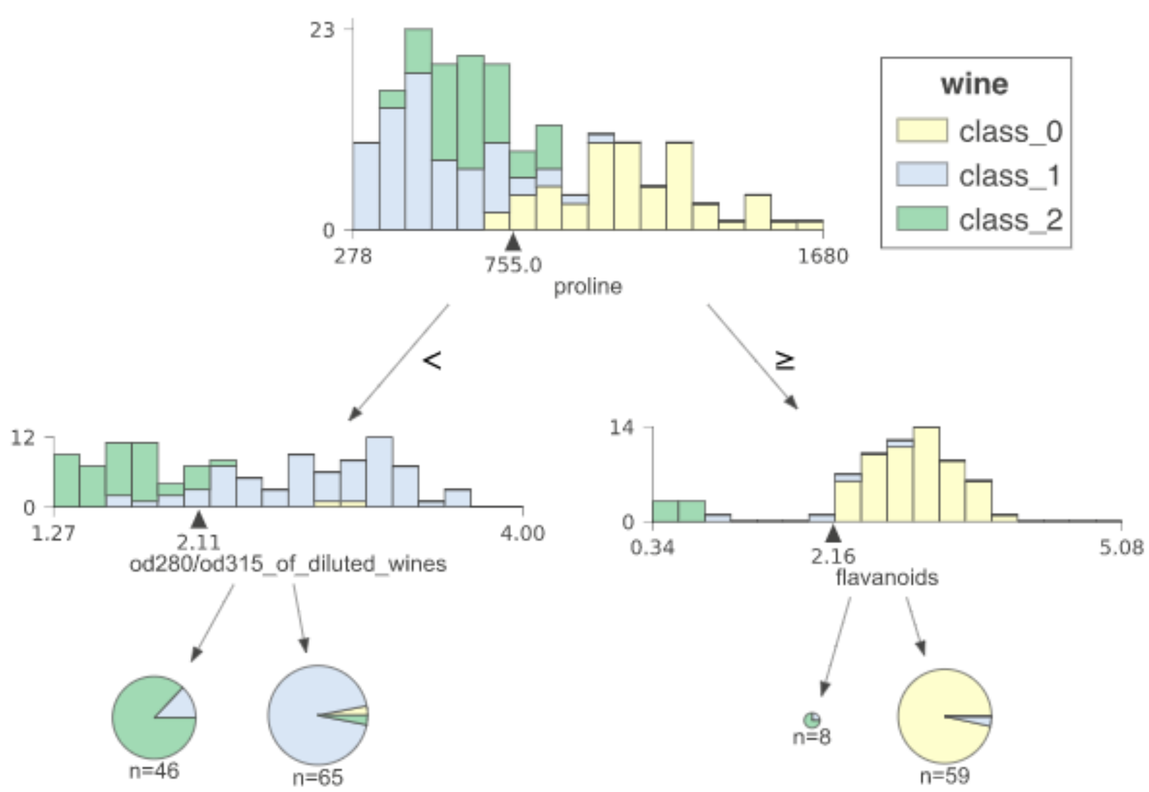


Sigmoid function σ(t) transforms a real number into a probability

Example of a decision tree for wine classification

Classify three types of wines:
classes 0, 1 and 2

Features:
chemical characteristics of a wine

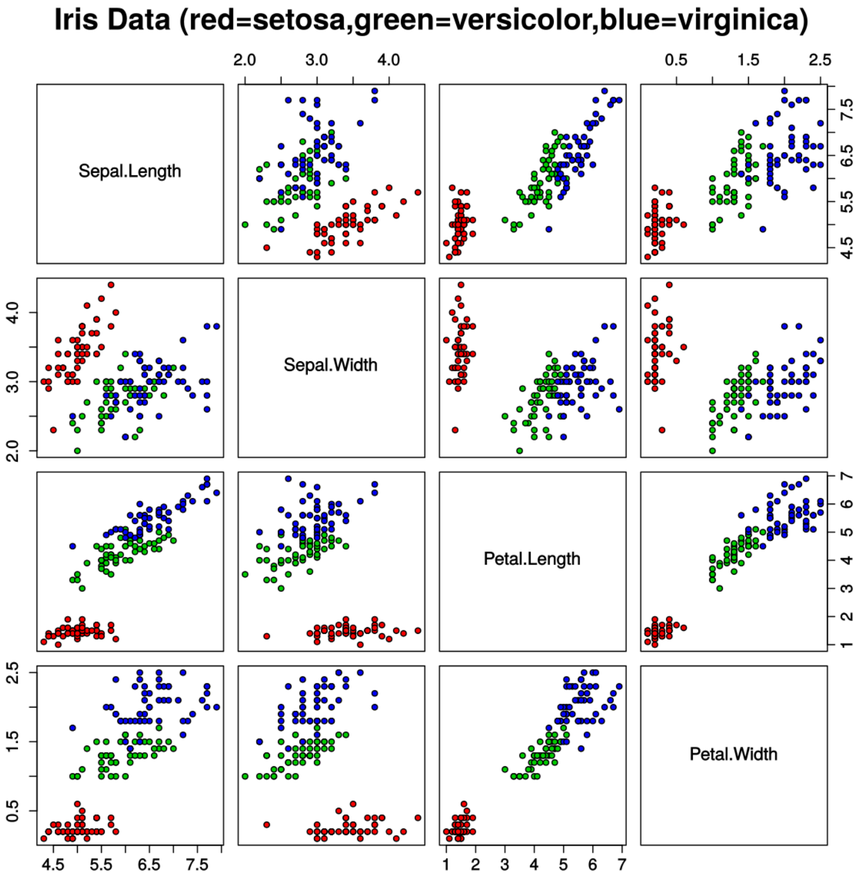


Example of a decision tree for iris classification

Classify three types of iris:
setosa, verscicolor и virginica

Features:
the length and width of the petal and the sepal

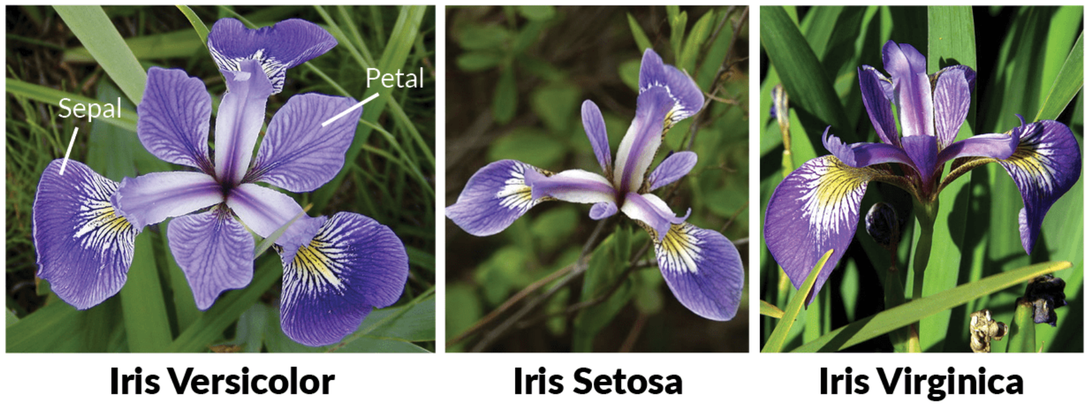
Iris Versicolor **Iris Setosa** **Iris Virginica**



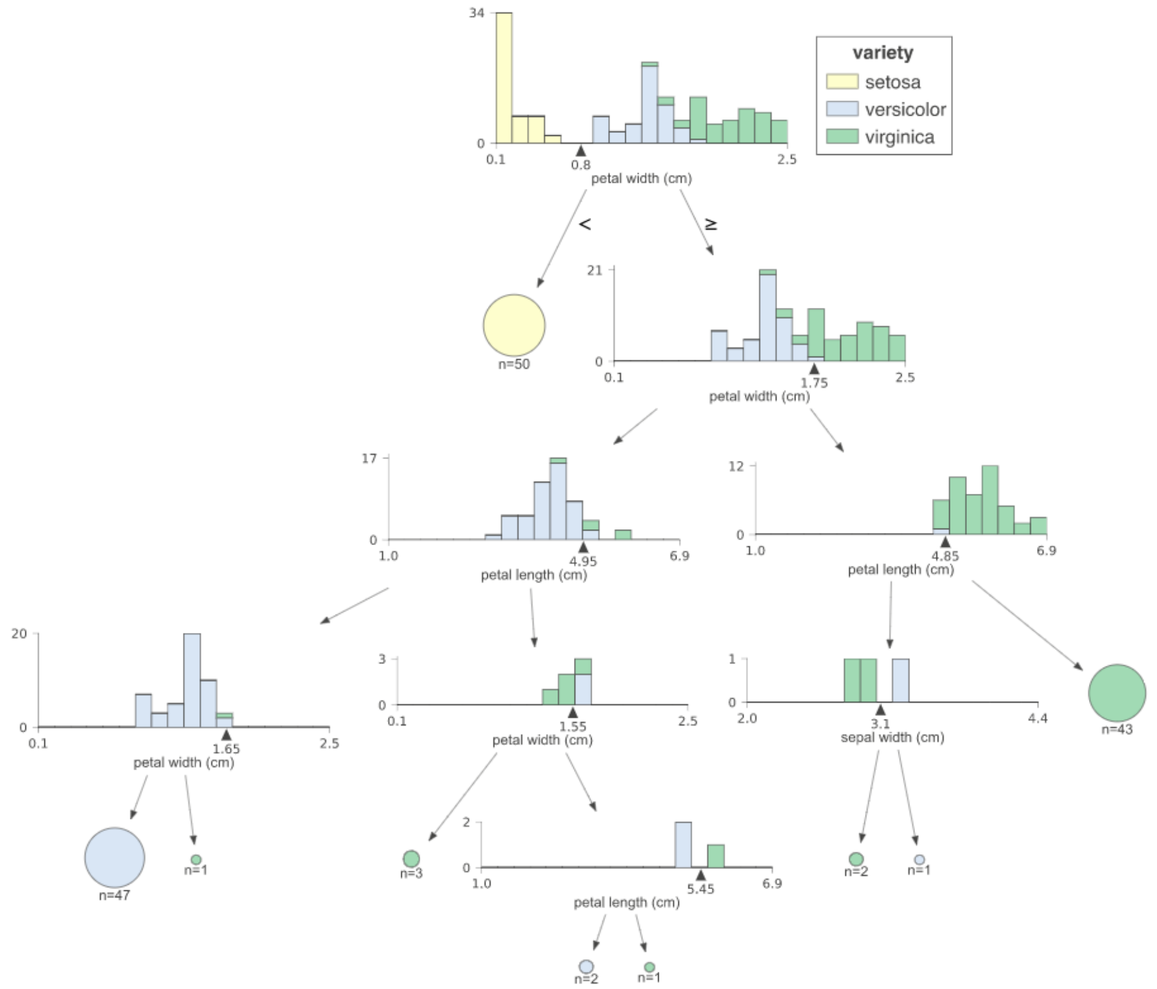
Example of a decision tree for iris classification

Classify three types of iris:
setosa, verscicolor и virginica

Features:
the length and width of the petal and the sepal

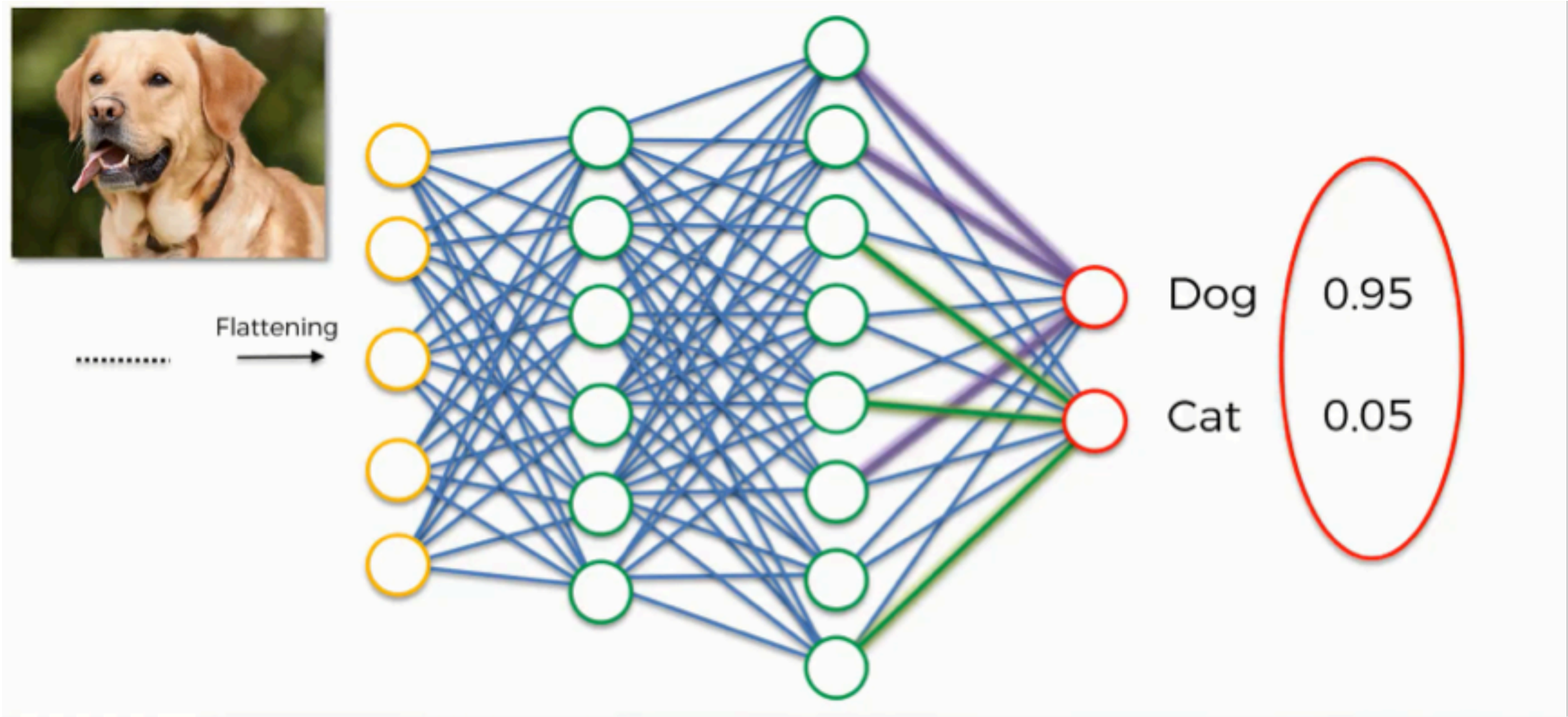


Iris Versicolor **Iris Setosa** **Iris Virginica**



http://www.r2d3.us/visual-intro-to-machine-learning-part-1/

Neural Networks



How to select a method for a particular problem

Start

> 50 of examples in a sample?

No

Yes

Table data?

No

Yes

Predict labels ?

Yes

> 1000 examples in a sample?

Yes

No

World of Neural Networks
In the next series…

The most used classification methods

Ensemble Methods:
RandomForestClassifier,
XGBClassifier

Logistic regression:
LogisticRegression