Introduction to Machine Learning

1. Definitions
   - what is machine learning?
   - supervised and unsupervised learning
2. Clustering
   - the Iris plants data
   - computing clusters by the k-means algorithm
3. Dimension Reduction
   - principal component analysis
   - training and testing

MCS 472 Lecture 22
Industrial Math & Computation
Jan Verschelde, 4 March 2026

Introduction to Machine Learning

1. Definitions
   - what is machine learning?
   - supervised and unsupervised learning
2. Clustering
   - the Iris plants data
   - computing clusters by the k-means algorithm
3. Dimension Reduction
   - principal component analysis
   - training and testing

What is Machine Learning?

There is not a clear boundary between statistics and machine learning

is a quote taken from Chapter 9 on *Machine Learning Basics*

in *Statistics with Julia. Fundamentals for Data Science, Machine Learning and Artificial Intelligence*,

a book by Yoni Nazarathy and Hayden Klok, Springer 2021.

learning from experience

A popular quote from computer scientist Tom Mitchell provides as a workable definition for machine learning.

Definition (learning from experience)

A *program can be said to learn* from experience E with respect to

- some class of tasks T and
- performance measure P,

if its performance at tasks in T, as measured by P, improves with experience E.

The quote above is taken from the following book:
Gavin Hackeling: *Mastering Machine Learning with scikit-learn. Apply effective learning algorithms to real-world problems using scikit-learn.* Packt publishing, 2014.

Introduction to Machine Learning

1. Definitions
   - what is machine learning?
   - supervised and unsupervised learning

2. Clustering
   - the Iris plants data
   - computing clusters by the k-means algorithm

3. Dimension Reduction
   - principal component analysis
   - training and testing

supervised and unsupervised learning

We distinguish between two types of machine learning:

- Supervised learning:
  given are labeled inputs and outputs,
  the program learns from examples of the right answers.

- Unsupervised learning:
  the program does not learn from labeled data.

Two most common supervised machine learning tasks
are classification and regression.

Examples of unsupervised learning
are clustering and dimensionality reduction.

finding distinctive elements in images

What are the distinctive elements in images?

Consider: *What Makes Paris Look Like Paris?* by Carl Doersch, Saurabh Singh, Abhinav Gupta, Josef Sivic, and Alexei A. Efros in *Communications of the ACM*, Vol 58, No 12, pages 103-110, 2015.

- Can we recognize the city from pictures taken in the city?
- We know where each picture was taken.

So this is an example of supervised learning.

Paris or London?

Captured from Doersch et al., *Comm. ACM*, 2015.

predicting the value of a house

What is the value of a house?

We are looking for a number.

- The value of a house is it sales price.
- If the house is not on the market, then there is no right answer.

So this is an example of unsupervised learning.

In practice, compare with recently sold houses in the neighborhood.

Introduction to Machine Learning

1. Definitions
   - what is machine learning?
   - supervised and unsupervised learning
2. Clustering
   - the Iris plants data
   - computing clusters by the k-means algorithm
3. Dimension Reduction
   - principal component analysis
   - training and testing

the Iris plants data

The Iris plants dataset, first used by Sir R. A. Fisher (1936),

- well known in the pattern recognition literature,
- used in statistics software such as R and
  in machine learning packages such as scikit-learn.

loading and examining the data

The Iris plants data is summarized below.

- There are 150 instances, partitioned into three classes. We have 50 instances of each class with names `"Iris-setosa"`, `"Iris-versicolor"`, and `"Iris-virginica"`.
- Each instance has four attributes: sepal length and width, petal length and width (measured in cm).
- We have features and labels:
  1. The attributes of each plant are called *features*, represented by a 4-by-150 floating point matrix.
  2. The names of the plants are called *labels*, represented by an array of 150 strings.

  We can view features as input X and labels as output Y.

the petal and sepal parts of a flower

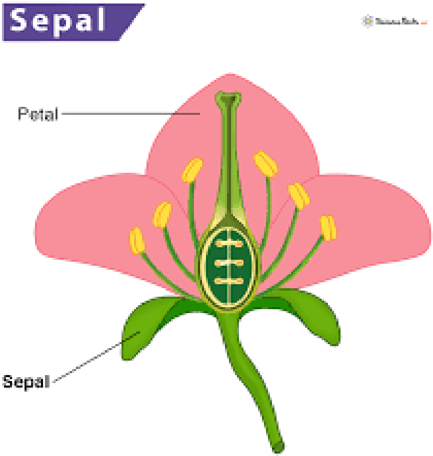


Image from `ScienceFacts.net`.

the Iris plants data in Julia

```
using MLDatasets: Iris

F, L = Iris(as_df=false)[:]
```

Then:

- The `F` is a 4-by-150 matrix of `Float64` elements.
- The `L` is a 150-element vector of elements of type `String`.

Introduction to Machine Learning

the k-means algorithm

Input: L a list of points;
k the number of centroids.

1. $C :=$ a random selection of k points of L;
2. repeat
   1. for each $p \in L$:
      if p is closest to C_i, then put p in the i-th cluster;
   2. recompute the centroids for each of the k clusters;

   until the clusters are stable.

Exercise 1: Write a Julia program for this algorithm.

Exercise 2: Discuss the cost of this algorithm.

the k-means algorithm in Julia

The k-means algorithm will be illustrated on the Iris plants data:

1. The features are in the 4-by-150 matrix `F`.
2. The labels are in the 150-element vector `L`.

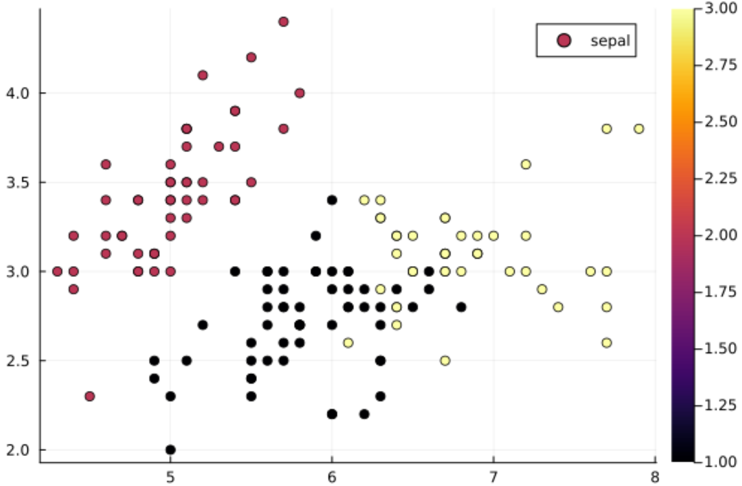
We use the `Clustering` package.

```
using Clustering
C = kmeans(F, 3)
dump(C)
```

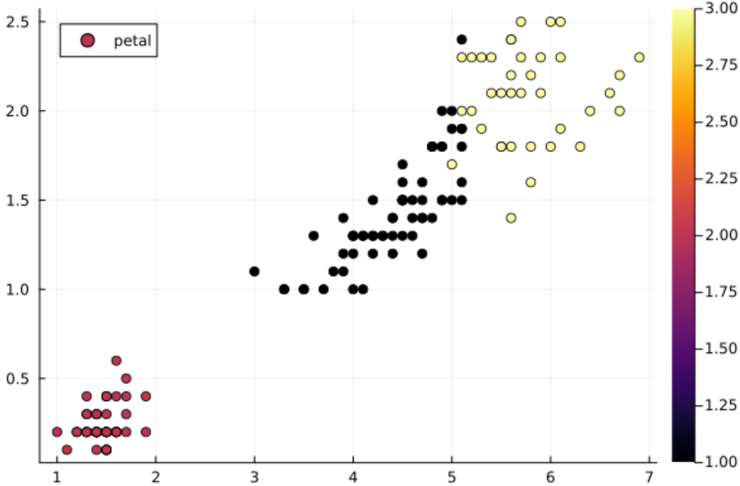
In the returned C.assignments we have 150 integers, 1, 2, or 3, corresponding to the three computed centroids in the data.

The `C.counts` are `[62, 50, 38]`, for one run.

assignments plotted versus sepal data



assignments plotted versus petal data



Introduction to Machine Learning

1. Definitions
   - what is machine learning?
   - supervised and unsupervised learning
2. Clustering
   - the Iris plants data
   - computing clusters by the k-means algorithm
3. Dimension Reduction
   - principal component analysis
   - training and testing

principal component analysis

Principal Component Analysis (PCA) derives an orthogonal projection to convert observations to linearly uncorrelated variables, called *principal components*.

Given a PCA model, we transform the observations x

$$y = P^T(x - \mu)$$

into its principal components y, with a projection matrix P.

The observations can be reconstructed from principal components:

$$\widetilde{x} = Py + \mu.$$

In Julia, the principal component analysis is in `MultivariateStats`.

Steps in the Principal Component Analysis

Three steps:

1. Standardize the data, via the formula

$$z = \frac{x - \text{mean}}{\text{standard deviation}},$$

   applied to each number x in the input.

2. Setup the covariance matrix C. For each pair (z_i, z_j),

$$\mathrm{cov}(z_i, z_j) \text{ is the } (i, j)\text{-th element of the matrix.}$$

   The matrix C is symmetric and positive definite.

3. Compute the eigenvectors and eigenvalues of C.
   Sort the eigenvectors in increasing magnitude of the eigenvalues.

Then the feature vector consist of the first p eigenvectors.

making a PCA model

The output of `M3d = fit(PCA, F; maxoutdim=3)` is

```
PCA(indim = 4, outdim = 3, principalratio = 0.99481691454981)

Pattern matrix (unstandardized loadings):
──────────────────────────────────────────
        PC1         PC2         PC3
──────────────────────────────────────────
1   0.743227   0.323137   -0.162808
2  -0.169099   0.359152    0.167129
3   1.76063   -0.0865096   0.0203228
4   0.737583  -0.0367692   0.153858
──────────────────────────────────────────

Importance of components:
──────────────────────────────────────────────────────────
                              PC1        PC2        PC3
──────────────────────────────────────────────────────────
SS Loadings (Eigenvalues)  4.22484   0.242244   0.0785239
Variance explained         0.924616  0.0530156  0.0171851
Cumulative variance        0.924616  0.977632   0.994817
Proportion explained       0.929434  0.0532918  0.0172747
Cumulative proportion      0.929434  0.982725   1.0
──────────────────────────────────────────────────────────
```

transforming the data for a plot

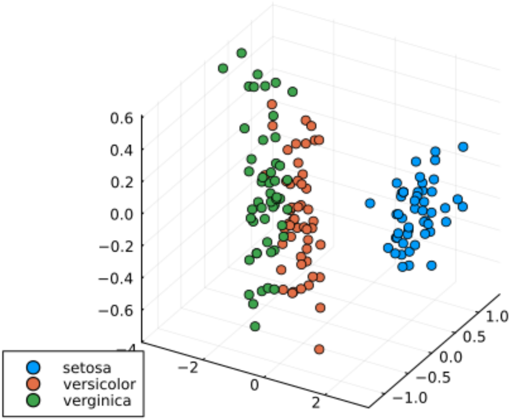
The PCA model `M3d` projects the 4-by-150 matrix `F` onto a three dimensional matrix of points:

```
F3d = transform(M3d, F)
```

The data was sorted according to the plant class, so we plot the groups of 50 as follows:

```
scatter(F3d[1,1:50], F3d[2,1:50], F3d[3,1:50],
        label="setosa")
scatter!(F3d[1,51:100], F3d[2,51:100],
         F3d[3,101:150], label="versicolor")
scatter!(F3d[1,101:150], F3d[2,101:150],
         F3d[3,101:150], label="verginica")
```

the data reduced to three dimensions



Introduction to Machine Learning

1. Definitions
   - what is machine learning?
   - supervised and unsupervised learning
2. Clustering
   - the Iris plants data
   - computing clusters by the k-means algorithm
3. Dimension Reduction
   - principal component analysis
   - training and testing

training and testing

Learning in a supervised setting is a 2-step process:

1. Use one part of the data set to train the model.
2. Test the model on the other part of the data set.

Applied to the Iris plants data:

1. Train a model with the odd indexed features.
2. Test the model on the even indexed features.

With principal component analysis:

1. Train the PCA model using the odd indexed features.
2. Transform the even indexed features and reconstruct.

Exercise 3:
Apply the training and testing to the Iris plants data
with the PCA model in `MultivariateStats`.

1. machine learning with scikit-learn

Gavin Hackeling: *Mastering Machine Learning with scikit-learn. Apply effective learning algorithms to real-world problems using scikit-learn.* Packt publishing, 2014.

One software that learns from experience is scikit-learn.

1. Read the book and the software documentation.
2. Describe how it fits in the computational ecosystem of Python.
3. Illustrate the capabilities by a good use case.
   How does machine learning predict the housing price?

2. computational geo-cultural modeling

Consider *What Makes Paris Look Like Paris?* by Carl Doersch, Saurabh Singh, Abhinav Gupta, Josef Sivic, and Alexei A. Efros in *Communications of the ACM*, Vol 58, No 12, pages 103-110, 2015.

This authors suggest *computational geo-cultural modeling* as the name for a new research area.

1. Write a summary of the paper.
2. What are the key algorithms needed for the computations?
3. Gather pictures from downtown and a Chicago suburb.
   What are the distinctive elements in the images?