Introduction
Machine Learning

Instructor: Polichshuk Yekaterina

Logistics

- **Instructor:** Polichshuk Yekaterina
  - Email: polichshuk.y.v@gmail.com
  - Office: 262

TA: Aidos Askhatuly

Email: aidos.askhatuly@gmail.com

Evaluation

Type of activity	Final scores
Assignments	26%
Participation	4%
Midterm	10%
Final Project Presentation	20%
Final exam	40%
Total	**100%**

Source Materials

- P. Harrington, Machine learning in Action(Recommended)
- T. Mitchell, ***Machine Learning***, McGraw-Hill
- Online courses:
- udacity.com - Introduction to machine learning

https://www.udacity.com/course/viewer#!/c-ud120/l-2254358555/e-3012748573/m-3035918544

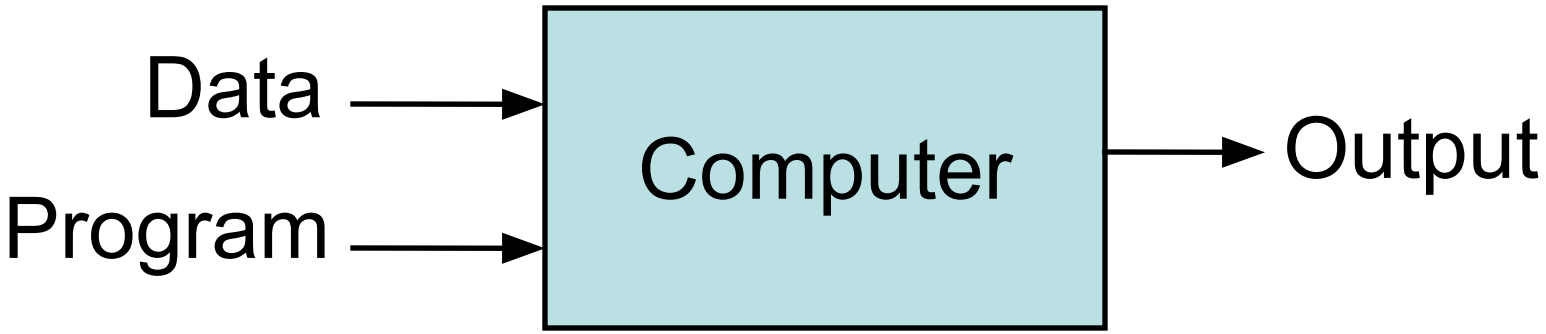
A Few Quotes

- “A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Chairman, Microsoft)
- “Machine learning is the next Internet”
  (Tony Tether, Director, DARPA)
- Machine learning is the hot new thing”
  (John Hennessy, President, Stanford)
- “Web rankings today are mostly a matter of machine learning” (Prabhakar Raghavan, Dir. Research, Yahoo)
- “Machine learning is going to result in a real revolution”
  (Greg Papadopoulos, CTO, Sun)
- “Machine learning is today’s discontinuity”
  (Jerry Yang, CEO, Yahoo)

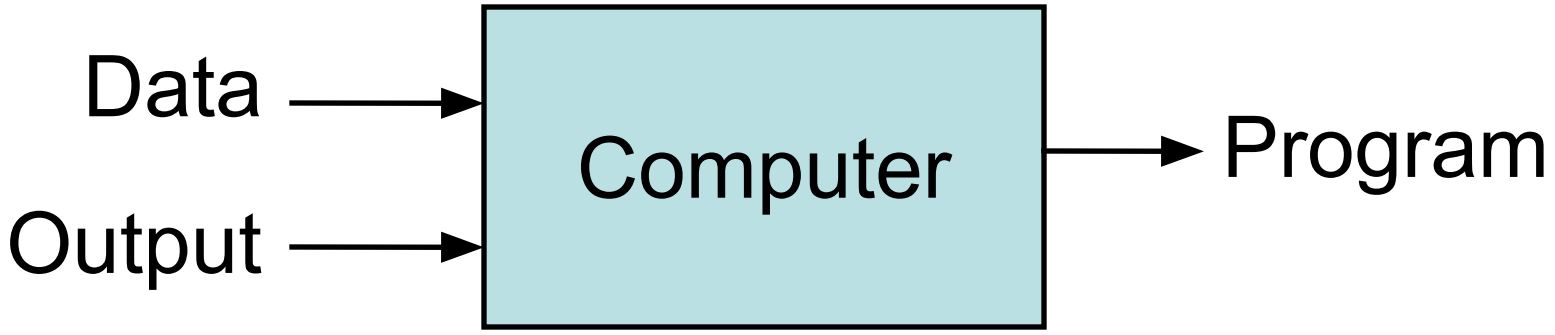
So What Is Machine Learning?

- Automating automation
- Getting computers to program themselves
- Writing software is the bottleneck
- Let the data do the work instead!

Traditional Programming



Machine Learning



Magic?

No, more like gardening

- **Seeds** = Algorithms
- **Nutrients** = Data
- **Gardener** = You
- **Plants** = Programs

Sample Applications

- Web search
- Computational biology
- Finance
- E-commerce
- Space exploration
- Robotics
- Information extraction
- Social networks
- Debugging
- [Your favorite area]

ML in a Nutshell

- Tens of thousands of machine learning algorithms
- Hundreds new every year
- Every machine learning algorithm has three components:
  - **Representation**
  - **Evaluation**
  - **Optimization**

Representation

- Decision trees
- Sets of rules / Logic programs
- Instances
- Graphical models (Bayes/Markov nets)
- Neural networks
- Support vector machines
- Model ensembles
- Etc.

Evaluation

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- K-L divergence
- Etc.

Optimization

- Combinatorial optimization
  - E.g.: Greedy search
- Convex optimization
  - E.g.: Gradient descent
- Constrained optimization
  - E.g.: Linear programming

Types of Learning

- **Supervised (inductive) learning**
  - Training data includes desired outputs
- **Unsupervised learning**
  - Training data does not include desired outputs
- **Semi-supervised learning**
  - Training data includes a few desired outputs
- **Reinforcement learning**
  - Rewards from sequence of actions

Inductive Learning

- **Given** examples of a function *(X, F(X))*
- **Predict** function *F(X)* for new examples *X*
  - Discrete *F(X)*: Classification
  - Continuous *F(X)*: Regression
  - *F(X)* = Probability(*X*): Probability estimation

What We'll Cover

- **Supervised learning**
  - Decision tree induction
  - Rule induction
  - Instance-based learning
  - Bayesian learning
  - Neural networks
  - Support vector machines
  - Model ensembles
  - Learning theory
- **Unsupervised learning**
  - Clustering
  - Dimensionality reduction

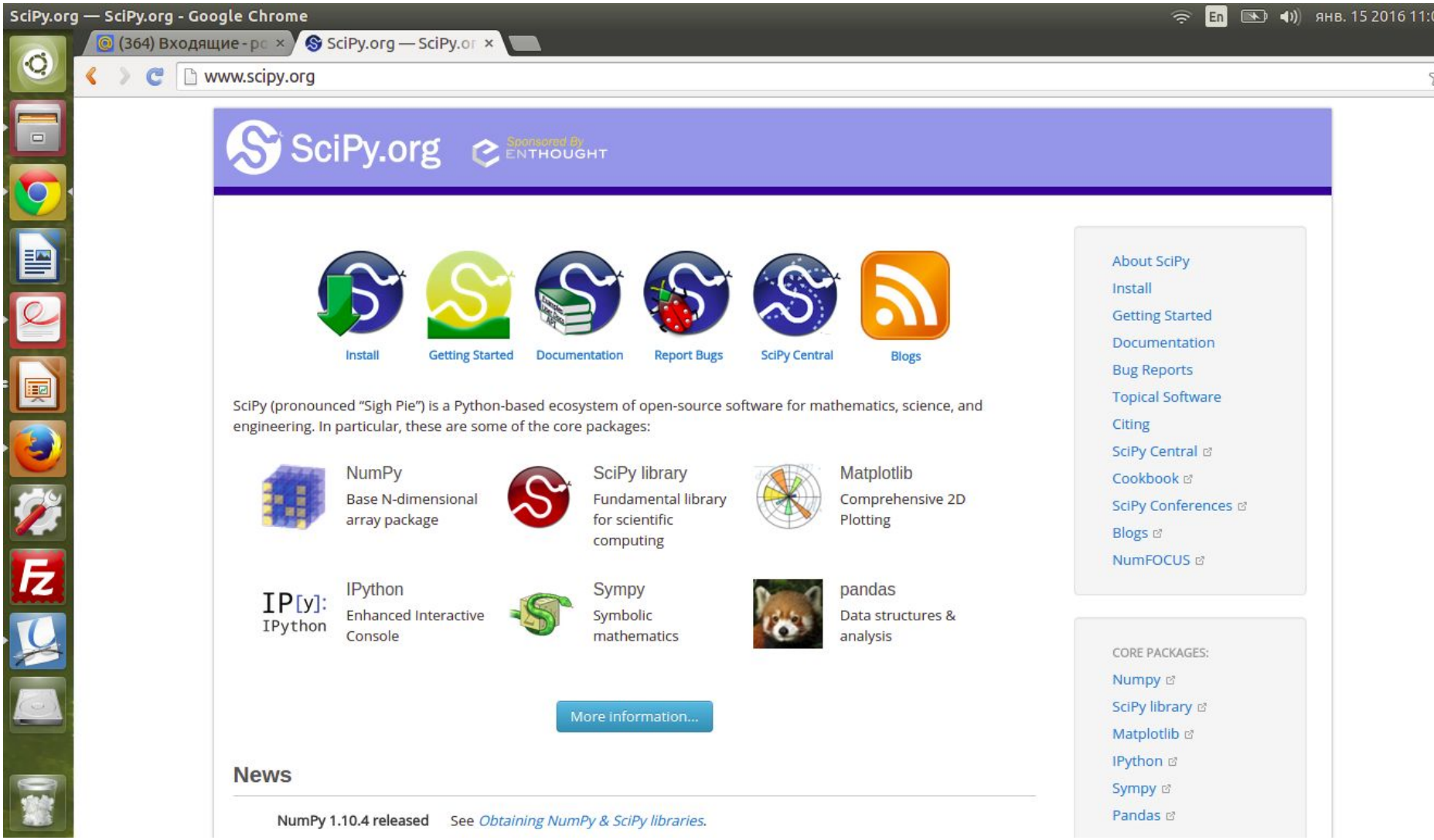
Steps in developing a machine learning application

- Collect data.
- Prepare the input data.
- Analyze the input data.
- Filter garbage
- Train the algorithm.
- Test the algorithm.
- Use it.

Programming languages

- Why Python?
- Python is a great language for machine learning for a large number of reasons.
- Python has clear syntax.
- it makes text manipulation extremely easy.
- A large number of people and organizations use Python, so there’s ample development and documentation.

Libraries: SciPy



Homework

- Read 1st chapter in “Machine learning in Action”
- Find any interesting material connect to ML