Fourth Industrial Summer School

Advanced Machine Learning

Introduction and Fundamentals

Session Objectives

- ✓ Introduction
- ✓ Fundamentals
- ✓ Regression exercise
- ✓ Summary

Machine Learning

- Arthur Samuel (1959): Field of study that gives computers the ability to learn without being explicitly programmed.
- Tom Mitchell (1998): Well-posed Learning Problem: A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

- Machine learning vs. algorithms?

Machine Learning Topics

- Unsupervised Learning
- Supervised Learning
  - KNNs
  - Linear and logistic regression
  - SVMs
  - Generative models
  - Decision trees and random forests
  - Artificial neural networks and deep learning
- Reinforcement Learning

Supervised Machine Learning

- Regression: Continuous output space
- Classification : Discrete output space

Prediction Models

- Types of Models
  - Discriminative models
  - Generative models
- Decision boundary
  - Linear decision boundaries
  - Non-linear decision boundaries

Linear Regression

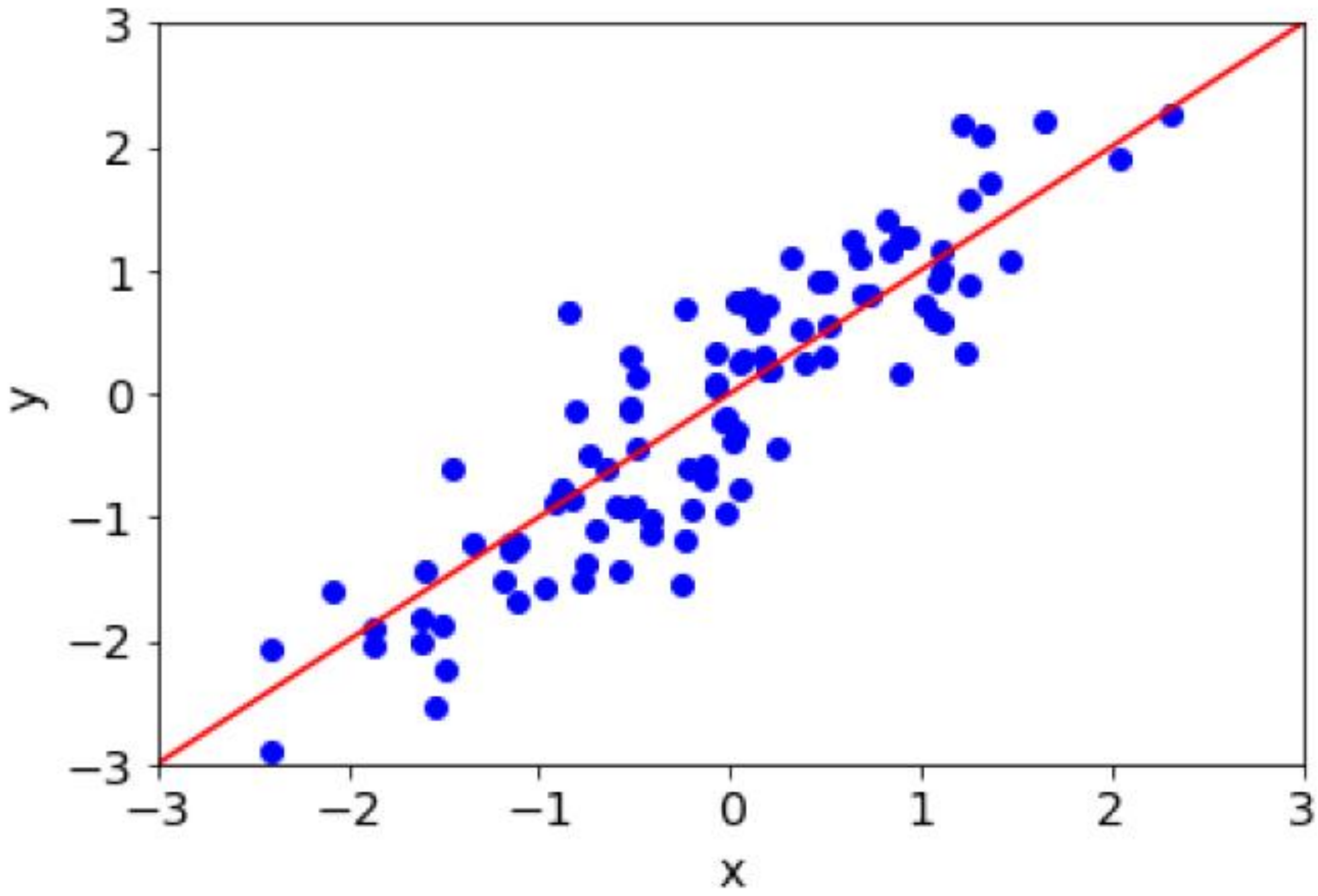
Regression problem:

- Response $y \in \mathbb{R}$
- predictor variables x ∈ $\mathbb{R}^{d}$

Solution:

- Given: data $(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)})$
- Return: Linear function of d variables (+ bias):
  - $f(x) = w_1 x_1 + w_2 x_2 + \cdots + w_d x_d + \mathrm{b}$
- Goal: minimize the loss function
- Penalize error using squared loss
  - $J(w, b) = \sum_{i=1}^{n} (y - (w.x + b))^2$

Linear Regression



Linear Regression: Closed-form solution

Then the loss function is minimized at:

- $w = (\mathrm{X}^T\mathrm{X})^{-1}(\mathrm{X}^T\mathbf{y})$
- Where, w are the weight including the bias term.

Linear Regression: Regularization

Regularization: Minimize squared loss plus a term that penalizes "complex" w:

Ridge:

- $L(w, b) = \sum_{i=1}^{n}(y - (w.x \; + \; b))^2 + \lambda||w||^2$
- Solution: $w = (\mathrm{X}^T\mathrm{X} + \lambda \mathrm{I})^{-1}(\mathrm{X}^T\mathrm{y})$

Lasso: tends to produce sparse w

- $L(w, b) = \sum_{i=1}^{n}(y - (w.x \; + \; b))^2 + \lambda||w||$

Classifier Training

- Loss function
- Minimizing loss (optimization problem)
- Distance measures
- Regularization

Loss Function

- Mean-squared error
- Cross-entropy loss

Minimizing Loss

Training the model

- Analytic solution vs. iterative solution
- Gradient descent and its variants

Distance Measures

- LP Norm
- Edit distance
- KL-Divergence (relative entropy)
- …

- When is a distance measure a metric?

Bias and Variance

- Train/dev/test
- Parameters vs. hyper-parameters

Case Study-Exercise

References

- Sanjoy Dasgupta, Machine Learning Fundamentals, UC San Diego
- Andrew Ng, Machine Learning, Stanford University
- Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, Foundations of Machine Learning, second edition, The MIT Press
- Andrew Ng, Machine Learning Yearning, deeplearning.ai