Machine Learing

Matteo Secco

February 27, 2021

Contents

1 Supervised Learning

input x also called features or attributes

output t also called targets or labels

Goal find a good approximation for $f : x \to t$

1.1 Tasks

Classification t is discrete

Regression t is continuous

Probability estimation t is a probability

1.2 When to use supervised learning?

- Human cannot perform the task
- Human can perform the task but cannot explain how
- Task changes over time
- Task is user-specific

1.3 Steps

To approximate f over the dataset $\mathcal{D}$

1. Define a loss function $\mathcal{L}$
2. Choose the hypothesis space $\mathcal{H}$
3. find $h \in \mathcal{H}$ that minimizes $\mathcal{L}$ over $\mathcal{D}$

1.4 Representation, Evaluation, Optimization

Examples of representation

- Linear models
- Instance-based
- Decision trees
- Set of rules
- Graphical models
- Neural networks

- Gaussian Processes
- Support vector machines
- Model ensembles

Examples of evaluation

- Accuracy
- Precision and recall
- Squared Error
- Likelihood
- Posterior probability
- Cost/Utility
- Margin
- Entropy
- KL divergence

Examples of optimization

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

1.5 Supervised learning taxonomy

- Parametric vs Nonparametric
  - Parametric: fixed and finite number of parameters
  - Nonparametric: the number of parameters depends on thetraining set
- Frequentist vs Bayesian
  - Frequentist: use probabilities to model the sampling process
  - Bayesian: use probability to model uncertainty about the estimate

- Empirical Risk Minimization vs Structural Risk Minimization
  - Empirical Risk: Error over the training set
  - Structural Risk: Balance training error with model complexity
- Direct vs Generative vs Discriminative
  - Generative: learns the joint probability distribution $p(x, t)$
  - Discriminative: learns the conditional probability distribution $p(t|x)$

1.6 Learning approaches

Direct approach Learn directly f from D

Discriminative approach

- Model $p(t|x)$
- Marginalize to find $E[t|x] = \int t \cdot p(t|x)\, dt$

Generative approach

- Model $p(x, t)$
- Infer $p(t|x)$ (Bayes rule)
- Marginalize to find $E[t|x] = \int t \cdot p(t|x)\, dt$