Artificial Intelligence

wu

April 27, 2019

Contents

1 Inference and Reasoning

1.1 Propositional logic

1.2 Predicate logic

1.3 First Order Inductive Learner

knowledge graph: node = entity, edge = relation. triplet (head entity, relation, tail entity)

2 Statistical learning and modeling

2.1 Machine Learning: the concept

2.1.1 Example and concept

Supervised learning problems applications in which the **training data** comprises examples of the input vectors along with their corresponding **target vectors** are known

classification and regression

Unsupervised learning problems the training data consists of a set of input vectors X **without any corresponding target values**

density estimation, clustering, hidden markov models

Reinforcement learning problem finding suitable actions to take in a given situation in order to maximize a reward. Here the learning algorithm is not given examples of optimal outputs, in contrast to supervised learning, but must instead discover them by a process of trial and error. A general feature of reinforcement learning is the trade-off between exploration and exploitation

types of machine learning

- supervised learning
  - classification: the output is categorical or nominal variable
  - regression: the output is read-valued variable
- unsupervised learning
- semi-supervised learning

- reinforcement learning
- deep learning

2.1.2 supervised learning: important concepts

- Data: labeled instances $< \boldsymbol{x}_i, \boldsymbol{y} >$
- features: attribute-value pairs which characterize each $\boldsymbol{x}$
- learning a discrete function: **classification**
- learning a continuous function: **regression**

Classification - A two-step process

- **model construction**
- **model usage**

2.2 example: polynomial curve fitting

2.3 probability theory review and notation

2.4 information theory

2.5 model selection

2.6 decision theory

3 Statistical learning and modeling - Supervised learning

3.1 Basic concepts

- **Linearly separable**
- **representation of class labels**
  - Two classes K = 2
  - K classes
    - 1-of-K coding scheme $\boldsymbol{t} = (0, 0, 1, 0, 0)^T$
  - Predict discrete class labels

* linear model prediction $y(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + w_0$ w: weight vector, w_0 bias/threshold
* nonlinear function $f(.) : R \to (0, 1)$
* generalized linear models $y(\boldsymbol{x}) = f(\boldsymbol{w}^T\boldsymbol{x} + w_0)$ f:activation function
* dicision surface $y(\boldsymbol{x})$ = constant $\to \boldsymbol{w}^T\boldsymbol{x} + w_0$ = constant

- **Three classification approaches**
  - discriminant function
    * least squares approach
    * fisher's linear discriminant
    * the perceptron algorithm of rosenblatt
  - use discriminant functions directly and don't compute probabilities

3.2 discriminant functions

3.2.1 Two classes

- Linear discriminant function $y(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + w_0$
  - Dicision surface $\Omega : y(\boldsymbol{x}) = 0$
  - the normal distant from the origin to the dicision surface $\frac{\boldsymbol{w}^T\boldsymbol{x}}{\|\boldsymbol{w}\|} = -\frac{w_0}{\|\boldsymbol{w}\|}$
  - if x_A, x_B lie on the decision surface $y(\boldsymbol{x}_A) = y(\boldsymbol{x}_B) = 0$, then $\boldsymbol{w}^T(\boldsymbol{x}_A - \boldsymbol{x}_B) = 0$. hence w is orthogonal to every vector lying within . $\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}$ is the normal vector of
  - $\boldsymbol{x} = \boldsymbol{x}_\perp + r\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}$ hence $r = \frac{y(\boldsymbol{x})}{\|blx\|}$. $y(\boldsymbol{x}_\perp) = 0 \to \boldsymbol{w}^T\boldsymbol{x} = -w_0 + r\frac{\boldsymbol{w}^T\boldsymbol{w}}{\|\boldsymbol{w}\|}$
  - $\tilde{\boldsymbol{w}} = (w_0, \boldsymbol{w}), \tilde{\boldsymbol{x}} = (x_0, \boldsymbol{x})$

3.2.2 K-class

- One-versus-the-rest classifier K - 1 classifiers each of which solves a two-class problem
- One-versus-one classifier K(K-1)/2 binary discriminant functions
- single K-class discriminant comprising K linear functions $y_k(\boldsymbol{x}) = \boldsymbol{w}_k^T\boldsymbol{x} + w_{k_0}$

- assigning a point x to class $\mathcal{C}_k$ if $y_k(\boldsymbol{x} > y_j(\boldsymbol{x}))$ for all jk
- dicision boundary between class $\mathcal{C}_k, \mathcal{C}_j$ is given $y_k(\boldsymbol{x}) = y_j(\boldsymbol{x}) \rightarrow (\boldsymbol{w}_k - \boldsymbol{w}_j)^T \boldsymbol{x} + (w_{k_0} - w_{j_0}) = 0$
- $\mathcal{R}_k$ is singly connected convex

3.2.3 Learning the parameters of linear discriminant functions

1. Least-squares approach
   - Problem
   - Learning
     - SSE function $SSE = \sum_{i=1}^{n}(y_i - f(x_i))^2 \; E_D(\widetilde{\boldsymbol{W}}) = 1/2\mathrm{Tr}\{(\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{W}} - \boldsymbol{T})^T(\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{W}} - \boldsymbol{T})\}$
2. fisher's linear discriminant from the view of dimensionality reduction
   $y \geq -w_0$ as class $\mathcal{C}_1$
   $m_1 = \frac{1}{N_1}\sum_{n\in\mathcal{C}_1} x_n, m_2 = \frac{1}{N_2}\sum_{n\in\mathcal{C}_2} x_n \xrightarrow{y=\boldsymbol{w}^T\boldsymbol{x}} m_2 - m_1 = \boldsymbol{w}^T(\boldsymbol{m}_2 - \boldsymbol{m}_1)$
3. the perceptron algorithm of rosenblatt

3.3 probalibilistic generative models

3.4 probabilistic discriminative models