Tutorial 7

Tutorial on kNN, SVM, MDP and Q-Learning

Luke Chang

The University of Auckland

May 2021

Topics

1. Bayesian Networks
2. Bayesian Networks
3. Bayesian Networks
4. K-Nearest Neighbour (kNN) Model
5. Support Vector Machine (SVM)
6. Markov Decision Process (MDP)
7. Q-Learning

Bayesian Networks

Example

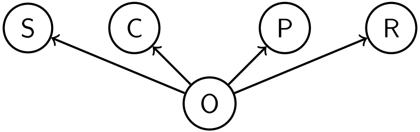
You are given a toxicity data set that describes chemical compounds with 5 *Boolean* attributes water solubility (**S**), cytochrominhibitor (**C**), contains phosphate (**P**), and cancerogenic in the rat model (**R**), and the outcome of some toxicity test (**O**). For the given dataset, could you learn a Bayesian network on the dataset?

S	**C**	**P**	**R**	**O**
TRUE	TRUE	FALSE	TRUE	Negative
TRUE	FALSE	TRUE	TRUE	Negative
FALSE	FALSE	TRUE	FALSE	Negative
FALSE	TRUE	TRUE	TRUE	Positive

If you condition on every attribute (join links top down), **O** will condition on $4! = 24$ possible combinations.

Bayesian Networks

S	**C**	**P**	**R**	**O**
TRUE	TRUE	FALSE	TRUE	Negative
TRUE	FALSE	TRUE	TRUE	Negative
FALSE	FALSE	TRUE	FALSE	Negative
FALSE	TRUE	TRUE	TRUE	Positive



P(O)	P(¬O)
0.25	0.75

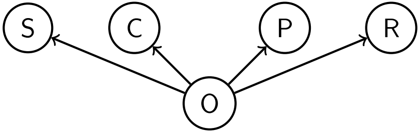
O	P(S)	P(¬S)
P	0	1.0
N	0.666	0.333

O	P(C)	P(¬C)
P	1.0	0
N	0.333	0.666

O	P(P)	P(¬P)
P	1.0	0.0
N	0.666	0.333

O	P(R)	P(¬R)
P	1.0	0.0
N	0.666	0.333

Bayesian Networks



P(O)	P(¬O)
0.25	0.75

O	P(S)	P(¬S)
P	0.0	1.0
N	0.666	0.333

O	P(C)	P(¬C)
P	1.0	0.0
N	0.333	0.666

O	P(P)	P(¬P)
P	1.0	0.0
N	0.666	0.333

O	P(R)	P(¬R)
P	1.0	0.0
N	0.666	0.333

A new instance with $S = T, C = F, P = F, R = F$, what is the probability of test positive?

$$\begin{aligned} P(O = P, S, \neg C, \neg P, \neg F) &= P(S|O)P(\neg C|O)P(\neg P|O)P(\neg F|O)P(O) \\ &= 0.25 \cdot 0.0 \cdot 0.0 \cdot 0.0 = 0.0 \end{aligned}$$

K-Nearest Neighbour (kNN) Model

The k-nearest neighbour fits for $\hat{Y}$ is defined as follows:

$$\hat{Y}(\mathbf{x}) = \frac{1}{k} \sum_{\mathbf{x} \in N_k(\mathbf{x})} y_i$$

where $N_k(\mathbf{x})$ is the neighbourhood of $\mathbf{x}$ defeined by the k closest points $\mathbf{x}$ in the training sample.

What does kNN do during training?

K-Nearest Neighbour (kNN) Model

The k-nearest neighbour fits for $\hat{Y}$ is defined as follows:

$$\hat{Y}(\mathbf{x}) = \frac{1}{k} \sum_{\mathbf{x} \in N_k(\mathbf{x})} y_i$$

where $N_k(\mathbf{x})$ is the neighbourhood of $\mathbf{x}$ defeined by the k closest points $\mathbf{x}$ in the training sample.

What does kNN do during training?

- Saving all training instances
- Algorithms used to compute the nearest neighbors:
  - Brute-force search
  - **KD Tree:** Splits from *median* on every feature; works well in lower dimensional data
  - **Ball Tree:** Also a binary tree which partitions data from N-dimensional hyper-sphere; the preferred method for high dimensional data

K-Nearest Neighbour (kNN) Model

What should you choose for the distance metric?

K-Nearest Neighbour (kNN) Model

What should you choose for the distance metric?

- Euclidean Distance: L_2-norm
- Manhattan Distance: L_1-norm, works better in higher dimensional data
- Mahalanobis Distance, Chebyshev Distance (L_∞-norm) and others

What are the limitations?

K-Nearest Neighbour (kNN) Model

What should you choose for the distance metric?

- Euclidean Distance: L_2-norm
- Manhattan Distance: L_1-norm, works better in higher dimensional data
- Mahalanobis Distance, Chebyshev Distance (L_∞-norm) and others

What are the limitations?

- Sensitive to noise
- Computational expensive at inference time (Scale by the size of training data)
- Does not scale well with larger dataset

Support Vector Machine (SVM)

Markov Decision Process (MDP)

Q-Learning