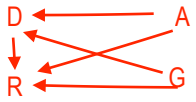
Quiz For Machine Learning: Models and Applications

Problem 1. A doctor gives a patient a (D)rug (drug or no drug) dependent on their (A)ge (old or young) and (G)ender (male or female). Whether or not the patient (R)ecovers (recovers or doesn't recover) depends on all D, A and G. In addition, A and G are assumed to be independent.

(a) Draw the belief network for the above case.



(b) Is it a Directed Acyclic Graph? Why? Yes

(c) Decompose the joint distribution $P(A, G, D, R)$ into four terms, each of which is a probability of conditional probability of only one variable. P(A,G,D,R) = P(A)*P(G)*P(D|A,G)*P(R|A,D,G)

(d) (True or False)

$P(D) = P(A = \text{“}Young\text{”}, D) + P(A = \text{“}Old\text{”}, D)$ T

$P(A|D = \text{“}Drug\text{”}) + P(A|D = \text{“}No\,Drug\text{”}) = 1$ F

$P(D = \text{“}Drug\text{”}|A) + P(D = \text{“}No\,Drug\text{”}|A) = 1$ T

Problem 2 (True or False)

- For a 1D Gaussian distribution, a small variance leads to a more flat bell shape. F
- For a 1D Gaussian distribution, a small mean leads to a more flat bell shape. F
- Rotation is a linear operator. T
- A vector space can have multiple sets of basis. T
- Orthogonal vectors are always linearly independent. T
- Making decision using posteriors can be treated as a special case of that using losses or risks. T
- Let λ_{ij} denote the loss incurred for taking action α_i when the true state is w_j. We have $\lambda_{ij} = \lambda_{ji}$. (2 points) F
- Let λ_{ij} denote the loss incurred for taking action α_i when the true state is w_j. We have $\lambda_{ii} > \lambda_{ji}$. (2 points) F

Problem 3 (True or False)

- One should always utilize part of the test samples to train a machine learning model. F
- Validation data should come from the test data. F
- If we assume a group of samples follow a Gaussian distribution, then the estimated mean using MLE equals to the mean of the samples. T
- The results of MAP and MLE differ given a uniform prior distribution. F
- Given samples from two categories, each of which follows a Gaussian distribution, the decision boundary should always be a linear function, for example a line or a plane. F

Problem 4 (True or False)

- As the model complexity increases, the training errors tend to decrease. T
- As the model complexity increases, the test errors tend to decrease. F
- MAP is a non-parametric approach for parameter estimation. F
- K-NN is a non-parametric approach. T
- Histogram estimation is a non-parametric method. T
- Histogram estimation requires no human-set parameters. F
- PCA aims to maximize the reconstruction errors by projecting higher-dimension data to a lower-dimension ones. F
- PCA can be treated as a subspece-selection approach. T
- 1-NN and 3-NN always yield the same result. F

Quiz 5

1) Gradient descent can be applied to any objective function (True of False). F

2) Gradient descent can always find the global optimal (True of False). F

3) Perception algorithm always converges (True of False). F

4) MSE can always find a decision hyperplane that correctly classifies the two classes, if they are linearly separable (True of False). F

5) SVM is an unsupervised learning method (True of False). F