Machine Learning. Fall 2013. Homework 1.

Due: Tuesday 30/April. 2013.

Either: Hand in hardcopy in class.

Or electronic to grader: Jianyu Wang, wjyouch@gmail.com

Question 1. A prize is hidden behind one of three doors A,B, and C. The contestant picks a door, say A, but it is left closed. The host opens door C and shows that there is no prize behind it. Should the contestant change his mind and select door B?

Formulate this problem as Bayes inference. What is the prior probability for the position of the prize before door C is opened? What is the posterior after door C has been opened? What should the contestant's decision rule be?

Question 2.

Describe the Bayes risk for making a binary decision $y \in \{-1, +1\}$. What is the Bayes decision (i.e. best decision to make) if the loss function is $l(y = 1, \hat{y} = 1) = l(y = -1, \hat{y} = -1) = 0$ and $l(y = 1, \hat{y} = -1) = 10$, $l(y = -1, \hat{y} = 1) = 1$? (Here $\hat{y}$ is the decision you make).

Suppose somebody tosses a fair (unbiased) coin and if the result is 'heads' you get nothing, otherwise you get 5 dollars. How much would you pay to play this game? What if you would win 500 dollars instead of 5 dollars?

Suppose you have vector-valued data $\{\vec{x}_i : i = 1, ..., N_1\}$ and $\{\vec{x}_i : i = 1, ..., N_{-1}\}$ from two classes $y \in \{-1, +1\}$. Describe how to learn Gaussian distributions for the distributions $P(\vec{x}|y = 1)$ and $P(\vec{x}|y = -1)$. What is the log-likelihood rule for classifying the data? Hence compute the decision rule. When is this decision rule the same as separating the data by a plane?

Question 3. Consider an exponential distribution $P(x|\lambda) = \frac{1}{Z[\lambda]} \exp\{\lambda \cdot \phi(x)\}$. Suppose you have data $\mathcal{X} = \{\S_\infty, \S_\in, ..., \S_\mathcal{N}\}$, how do you estimate the parameters

λ of the distribution using maximum likelihood?

The Bernoulli distribution has form $P(x) = \theta^x (1-\theta)^{1-x}$, where $x \in \{0, 1\}$ (i.e. x takes value 0 or 1) and $\theta \in [0, 1]$. Re-express the Bernoulli distribution as an exponential distribution.

Describe how to estimate the parameter θ by maximum likelihood from $\mathcal{X}$.

Question 4. What is the entropy of the exponential distribution $P(x|\lambda) = \frac{1}{Z[\lambda]} \exp\{\lambda \cdot \phi(x)\}$?

What is the entropy of the Bernoulli distribution? Derive the Bernoulli distribution by using the maximum entropy principle.

Question 5.

Consider a hypothesis space H of classifiers and a dataset of N points in general position. There are 2^N ways to label these points as either positive or negative. What does it mean to say that H shatters the points? What is the Vapnik-Chervonenkis (VC) dimension of H?

Explain the difference between generalization and memorization. Can you generalize if the VC dimension of your hypothesis space is larger than the number of datapoints?

In class we showed that the VC dimension of planes/lines in two dimensions was three – e.g., if we have three points in two dimensions which are in general position (i.e. they do not lie on a line) then we always find a classifier which gets perfect classification. Now suppose the hypothesis space is the space of rectangles aligned to the axes in two-dimensional space (all points within the rectangle are positive and outside negative, or all points outside are positive and inside are negative). What is the VC dimension? What is the VC dimension if the hypothesis space is the set of ellipses?

Question 6.

Alpaydin. Chp (4). Question 7. Use Matlab or equivalent.