Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur
End-semester Examination, Autumn 2015
CS60050: Machine Learning

Full Marks: 100 **Time: 3 Hrs**

Answer ALL questions. You can use calculators.

Q1. Which of the following is true or false. Explain with reasons (max 2-3 sentences):

a. Rule post pruning in C4.5 can sometimes produce rules which do not correspond to a decision tree. **[5 marks]**

b. SMO algorithm can be used to solve any quadratic programming problem. **[5 marks]**

c. In SGD, the objective function value decreases monotonically with the iterations. **[5 marks]**

d. All conditional independences in encoded in the Bayesian network are preserved in the MRF obtained after moralization. **[5 marks]**

e. The perceptron learning algorithm is a special case SGD. **[5 marks]**

f. The Nesterov and Vial's proof for convergence of SGD applies to perceptron algorithm as well. **[5 marks]**

Q2. Answer the following:

a. Write a stochastic gradient descent algorithm, for optimizing a function $L(\theta) = {}^{1}/_{N} \sum_{n=1}^{N} l(x_n, y_n, \theta)$. **[5 marks]**

b. Define the expected loss function, clearly mentioning the random variables. Write the formula for convergence in expectation, clearly defining each variable. **[5 marks]**

c. Derive a condition on the step sizes which guarantees convergence in expectation. **[5 marks]**

d. Consider a situation where a biased coin with probability of head (p) is tossed N times and the outcomes ($\text{head} = 1, \text{tail} = 0$) are recorded in random variables, $x_n, n = 1, \ldots, N$. Derive the log likelihood function for estimating parameter, p. Derive an SGD algorithm for solving the Max-likelihood problem derived above. **[10 marks]**

Q3. Consider the following dataset for predicting **outcome** of a tennis match between Federer and Nadale:

Time	**Match type**	**Court surface**	**Best Effort**	**Outcome**
Morning	Master	Grass	1	F
Afternoon	Grand slam	Clay	1	F
Night	Friendly	Hard	0	F
Afternoon	Friendly	Mixed	0	N
Afternoon	Master	Clay	1	N
Afternoon	Grand slam	Grass	1	F
Afternoon	Grand slam	Hard	1	F
Afternoon	Grand slam	Hard	1	F
Morning	Master	Grass	1	F
Afternoon	Grand slam	Clay	1	N
Night	Friendly	Hard	0	F
Night	Master	Mixed	1	N
Afternoon	Master	Clay	1	N
Afternoon	Master	Grass	1	F
Afternoon	Grand slam	Hard	1	F
Afternoon	Grand slam	Clay	1	F

a. Write the formula for information gain for an attribute. Compute information gain for all attributes. Which is selected as a root node? **[10 marks]**

b. Compute the structure of the decision tree computed by ID3 algorithm. **[5 marks]**

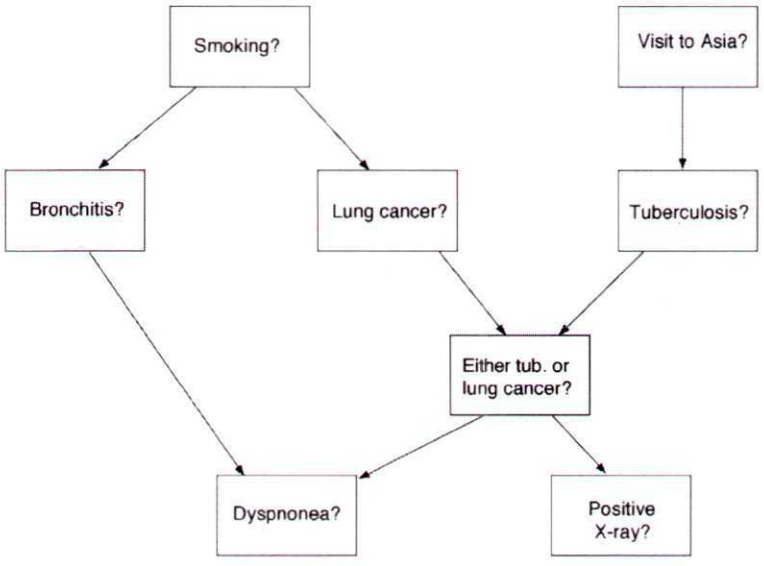
Q4. Show 5 updates of perceptron algorithm for the following data, starting at $w = (0,0)$:

{(-,(1,-4)), (+1,(5,-1)), (+,(3,3)), (+,(-1,5)), (-,(-5,2)), (-,(-2,-2))} **[10 marks]**

Q5.

a. Define d-separation in Bayesian Networks stating all the conditions. How does d-separation help in determining conditional independence? **[5 marks]**

b. Consider the following Bayesian network connection various factors related to chest diseases: **[10 marks]**

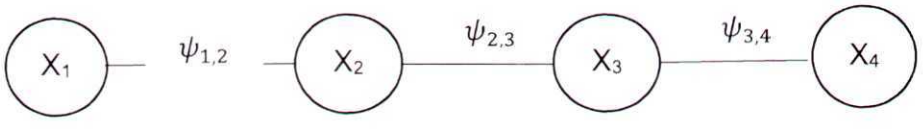


i. Are Bronchitis and Tuberculosis independent when nothing is observed?

ii. Are they independent after observing x-ray?

iii. Are they independent after observing both x-ray and smoking?

Q6. Consider the following Markov Random Field with the corresponding potentials:



$(X_i, X_{i+1}) = 1 - X_i X_{i+1}$ for odd i, and $\psi_{i,i+1}(X_i, X_{i+1}) = 1 + X_i X_{i+1}$ for even i; where $X_i \in \{+1, -1\}$. Find $P(X_2)$ and $x_3^* = \max_{x_3} P(X_3|X_2 = 1)$ using an appropriate message-passing algorithm.

[20 marks]