

Examen Final de Aprendizaje

(11 de enero de 2011)

Duración: 2 horas

1. (1.5 puntos) Supongamos que tenemos un lenguaje de descripción de conceptos que admite solo formulas conjuntivas puras y tenemos el dominio de las películas con los siguientes atributos:

Sonoro:	si, no
Color:	si, no
Nacionalidad:	USA, España, Francia, Italia
Duración:	cortometraje, medimetraje, largometraje
Género:	comedia, drama, acción, negro, policiaca, aventuras

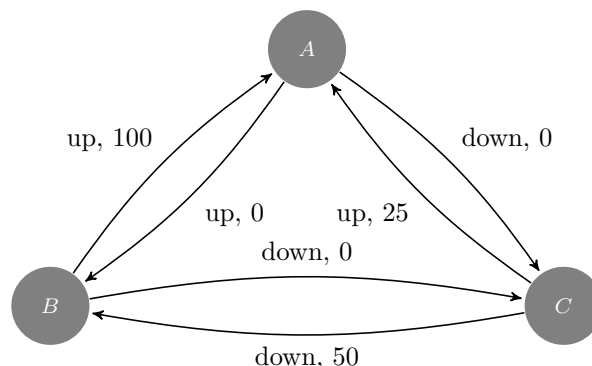
Genera el conjunto mínimo de ejemplos para aprender el siguiente concepto {comedias americanas mudas} y muestra la evolución del método del espacio de versiones para aprenderlo.

2. (1 punto) En la siguiente tabla está representada las distribuciones de probabilidad aprendidas mediante naïve bayes a partir de un conjunto de ejemplos pertenecientes a tres clases. El número de ejemplos de cada clase utilizados para general la tabla son: clase 1 = 300 ejemplos, clase 2 = 200 ejemplos, clase 3= 100 ejemplos. El conjunto de datos consta de tres atributos A1, A2, A3, con los siguientes valores $A1=\{a,b\}$, $A2=\{R,G,B\}$, $A3=\{x,y,z\}$.

	A1		A2			A3		
	a	b	R	G	B	x	y	z
Clase 1	0.3	0.7	0.1	0.3	0.6	0.1	0.6	0.3
Clase 2	0.9	0.1	0.4	0.5	0.1	0.3	0.3	0.4
Clase 3	0.6	0.4	0.7	0.1	0.2	0.4	0.4	0.2

Dado el siguiente ejemplo $ej1= (A1=a,A2=G,A3=y)$, ¿que probabilidad de pertenencia a cada clase le asignaría el clasificador aprendido?

3. (1.5 puntos) Queremos construir un sistema capaz de controlar un proceso con tres estados {A, B, C} donde se pueden realizar las acciones up y down. El siguiente gráfico indica como es la función de cambio de estado (δ) y el refuerzo (r) obtenido por cada acción:



Supondremos que para entrenar el sistema aprenderemos a partir de tres secuencias de entrenamiento:

- Estado Inicial=A, acciones={up,up,up}
- Estado Inicial=B, acciones={down,up,down}

c) Estado Inicial=C, acciones={down,down,up}

Calcula la función \hat{Q} que obtendríamos después de aprender estas tres secuencias mediante el algoritmo de Q-Learning, considerando que el valor del parámetro γ es 0.9. Muestra la evolución de la función \hat{Q} paso a paso.

4. (1.5 puntos) **Árboles de decisión**

- a) Los árboles de decisión buscan el árbol mas pequeño que es capaz de clasificar un conjunto de ejemplos ¿Por qué este tipo de árboles son mejores? ¿Que desventaja tendría un árbol más grande igual de predictivo?
- b) ¿Como es posible que el error estimado de un árbol de decisión disminuya tras el proceso de pruning, si estamos eliminando ejemplos del árbol?

5. (1.5 puntos) **K-nn**

Deseamos utilizar el algoritmo K-nearest neighbour para construir un clasificador para una aplicación concreta.

- a) ¿Cómo influye el número de ejemplos en la calidad de la predicción? Justifícalo.
- b) ¿Cómo influye el número de ejemplos en el tiempo de aprendizaje? Justifícalo.
- c) ¿Cómo influye el número de ejemplos en el tiempo de predicción? Justifícalo.
- d) ¿Cómo influye el número de atributos en la calidad de la predicción? Justifícalo.

6. (1.5 puntos) **Bayes**

- a) El algoritmo de naive bayes aproxima los conceptos mediante una distribución de probabilidad conjunta ¿qué supone este algoritmo sobre esta distribución de probabilidad? Las redes bayesianas también pueden usarse como clasificadores, ¿que diferencia existe entre la distribución de probabilidad que aprendería una red bayesiana y la que aprende el naive bayes?
- b) ¿Cómo crees que le afecta el ruido en los ejemplos de entrenamiento al algoritmo de naive bayes? ¿y los atributos no relevantes?

7. (1.5 puntos) **No supervisado**

- a) Explica el funcionamiento de los algoritmos *K-means* y *Expectation Maximization*. Comenta en que se parece uno al otro y que relación tiene el algoritmo *Expectation Maximization* con el aprendizaje bayesiano.

Las notas se publicarán el **17 de enero**.

Machine Learning Final Exam

(January 11th 2011)

Time: 2 hours

1. (1.5 points) Assume that you have a concept description language that only allows pure conjunctive formulas and we have the movies domain with the following attributes:

Sound:	yes, no
Color:	yes, no
Country:	USA, Spain, France, Italy
Duration:	long, medium, short
Genre:	comedy, drama, action, noir, thriller, adventure

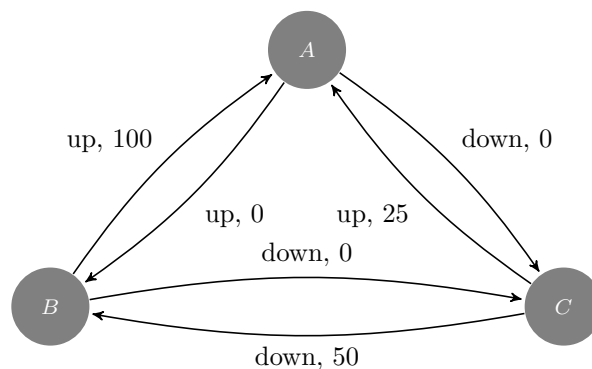
Find the minimum set of examples to learn the concept {silent american comedies} and show the evolution of the learning process obtained by the version space algorithm.

2. (1 point) The following table represents the probability distributions learned using the naive bayes algorithm using a set of examples from three classes. The number of examples of each class used for learning this classifier are: class 1 = 300 examples, class 2 = 200 examples, class 3 = 100 examples). This domain has three attributes A1, A2, A3, with the following values A1={a,b}, A2={R,G,B}, A3={x,y,z}.

	A1		A2			A3		
	a	b	R	G	B	x	y	z
Class 1	0.3	0.7	0.1	0.3	0.6	0.1	0.6	0.3
Class 2	0.9	0.1	0.4	0.5	0.1	0.3	0.3	0.4
Class 3	0.6	0.4	0.7	0.1	0.2	0.4	0.4	0.2

Given the example $ex_1 = (A1=a, A2=G, A3=y)$, what are the probabilities that the naive bayes classifier will assign to this example for each class?

3. (1.5 puntos) We want to build a system able to control a process with three states {A, B, C} where we can perform the actions up and down. The following figure shows the state transition function (δ) and the reinforcement (r) obtained by each action:



We will assume that in order to train the model we are going to use the following training sequences:

- Initial state=A, actions={up,up,up}
- Initial state=B, actions={down,up,down}

c) Initial state=C, actions={down,down,up}

Compute the \hat{Q} function that is obtained after processing these three sequences using the Q-Learning algorithm using 0.9 as the value of the parameter γ . Show the evolution of the \hat{Q} function step by step.

4. (1.5 points) **Decision trees**

- a) Decision trees algorithms look for the smallest tree able to classify a set of examples. Why this kind of trees are better than others? Why a larger tree equally predictive is not a good choice?
- b) How is it possible that the estimated error of a decision tree is lower after the pruning process when with this process we are discarding examples from the dataset?

5. (1.5 points) **K-nn**

We want to use the K-nearest neighbour algorithm to build a classifier for a problem.

- a) What is the influence of the number of examples on the quality of the predictions? Justify your answer.
- b) What is the influence of the number of examples on the learning time? Justify your answer.
- c) What is the influence of the number of examples on the prediction time? Justify your answer.
- d) What is the influence of the number of attributes on the quality of the predictions? Justify your answer.

6. (1.5 points) **Bayes**

- a) Naive bayes learns concepts by estimating the joint probability distribution of the classes from the dataset. What is the assumption that this algorithm makes about this probability distribution? Bayesian networks can also be used as classifiers, what is the difference between the probability distribution that learns a bayesian network and the one that learns naive bayes?
- b) For the Naive Bayes algorithm, how affects noise in the training data to the quality of the model learned using this algorithm? and irrelevant attributes? Elaborate your answer.

7. (1.5 points) **Unsupervised**

- a) Explain the *K-means* and *Expectation Maximization* algorithms. Comment the similarities among this two algorithm and the relation of the *Expectation Maximization* algorithm to bayesian learning.

The grades will be published **january 17th**.