

Apellidos : .....

Nombre : .....

Sólo se corregirán los ejercicios escritos a **bolígrafo azul** o **bolígrafo negro**

El siguiente ejercicio aborda el problema del etiquetado sintáctico. Para simplificar el problema vamos a reducir las categorías sintácticas a las siguientes 5 posibles:

- **Ifr**: Inicio de frase
- **Det**: Determinante
- **Nom**: Nombre
- **Ver**: Verbo
- **Pre**: Preposición

Consideremos la siguiente frase:

Juan	para	el	coche
------	------	----	-------

siguiendo el modelo de etiquetas propuesto se representará como:

Ifr	Nom	Ver	Det	Nom
-----	-----	-----	-----	-----

Es decir, introduciremos una etiqueta para indicar el inicio de cada frase, y cada palabra se etiquetará con la etiqueta sintáctica correspondiente (de nuevo, para simplificar no consideraremos ambigüedad léxica). En el ejemplo anterior hemos indicado que la palabra *Juan* es de tipo **Nom**, y así sucesivamente.

Consideremos ahora el siguiente corpus ya etiquetado

Ifr	Nom	Ver	Det	Nom		
Ifr	Det	Nom	Ver	Pre	Det	Nom
Ifr	Det	Nom	Pre	Det	Nom	Ver

### Ejercicio 1. [1,5 puntos]

Describe la noción de bigrama suavizado, y construye el bigrama suavizado correspondiente para el corpus de tres ejemplos anteriores.

A continuación, nos dan la siguiente frase:

Juan	para	para	dormir
------	------	------	--------

de donde sabemos que tanto *Juan* como *coche* son de categoría **Nom**, pero la palabra *para* puede ser tanto **Pre** o **Ver**, es decir, puede ser una preposición o un verbo.

Por tanto, obtenemos cuatro interpretaciones posibles. Indica cuales son estas cuatro posibilidades y a continuación calcula cuál de las interpretaciones maximiza el cálculo de probabilidades usando los bigramas previamente construidos. Indicar el modelo matemático utilizado y justificar su utilización.

**Ejercicio 2. [1,5 puntos]**

En este ejercicio se debe construir un perceptrón que nos ayude con el problema del etiquetado sintáctico.

En primer lugar, se debe definir en qué consiste un perceptrón, y en concreto el algoritmo de entrenamiento basado en umbral.

En segundo lugar diseñar un perceptrón que utilizará como entrada dos categorías sintácticas, y devolverá como resultado si la siguiente palabra sería de tipo nombre o no. Por ejemplo, si suponemos que tenemos una frase que comienza por **Ifr Det** queremos saber si la siguiente palabra será **Nom** o no. Es importante tener en cuenta que las entradas son variables (podrán ser cualquiera de las categorías contempladas), pero la salida debe indicar siempre si es o no un nombre.

El ejercicio consistirá en construir el modelo del perceptrón, definiendo los elementos que lo constituyen, a continuación especificar la lista de ejemplos de entrenamiento a partir del corpus indicado, y en tercer lugar hacer una simulación del entrenamiento. Esta simulación debe utilizar pesos iniciales 0,1 y un factor de aprendizaje igual a 0,2. Para reducir los cálculos es suficiente con aplicar el algoritmo a los ejemplos de entrenamiento generados por la primera frase del corpus. Se debe indicar cuales son estos ejemplos y cuales son los valores de los pesos al terminar este entrenamiento.