

Trabajo.3: Preguntas de Teoría

Fecha de entrega: 27 Mayo 2017. Valor máximo: 10 puntos

NORMAS DE DESARROLLO Y ENTREGA DE TRABAJOS

Para este trabajo como para los demás es obligatorio presentar un informe escrito (hacerlo en pdf, MS Word). **Sin este informe se considera que el trabajo NO ha sido presentado.**

Normas para el desarrollo de los Trabajos: EL INCUMPLIMIENTO DE LAS NORMAS (*) SIGNIFICA PERDIDA DE 2 PUNTOS POR CADA INCUMPLIMIENTO.

- En su informe de contestación debe incluir todas las preguntas en el orden y tal y como se les formula en este documento. (*)
- Las contestaciones irán a continuación de cada pregunta, dejando en blanco las que no conteste. (*)
- Todas las contestaciones deben ser justificadas con argumentos. Sin argumentos la pregunta se considera no contestada.
- Todas las justificaciones matemáticas deben contener todos y cada uno de los pasos de la misma. En caso de duda la contestación no se considerará válida.
- **Forma de entrega:** Subir el pdf a la web de Turnitin.

Todas las preguntas tienen el mismo valor

1. Considere un modelo de red neuronal con dos capas totalmente conectadas: d unidades de entrada, n_H unidades ocultas y c unidades de salida. Considere la función de error definida por $J(\mathbf{w}) \equiv \frac{1}{2} \sum_{k=1}^c (t_k - c_k)^2 = \frac{1}{2} \|\mathbf{t} - \mathbf{z}\|^2$, donde el vector \mathbf{t} representa los valores de la etiqueta, \mathbf{z} los valores calculados por la red y \mathbf{w} los pesos de la red. Considere que las entradas a la segunda capa se calculan como $z_k = \sum_{j=0}^{N_H} y_j w_{kj} = \mathbf{w}_k^t \mathbf{y}$ donde el vector \mathbf{y} representa la salida de la capa oculta.
 - a) Deducir con todo detalle la regla de adaptación de los pesos entre la capa oculta y la salida.
 - b) Deducir con todo detalle la regla de adaptación de los pesos entre la capa de entrada y la capa oculta.Usar θ para notar la función de activación.
2. Tanto “bagging” como validación-cruzada cuando se aplican sobre una muestra de datos nos permiten dar una estimación del error de un modelo ajustado a partir de dicha muestra de datos. Discuta cuál de los dos métodos considera que obtendrá una mejor estimación del error. Especificar con precisión las razones.
3. Considere que dispone de un conjunto de datos linealmente separable. Recuerde que una vez establecido un orden sobre los datos, el algoritmo perceptron encuentra un hiperplano separador iterando sobre los datos y adaptando los pesos de acuerdo al algoritmo

Algorithm 1 Perceptron

```
1: Entradas:  $(\mathbf{x}_i, y_i), i = 1, \dots, n$ ,  $\mathbf{w} = 0$ ,  $k = 0$ 
2: repeat
3:    $k \leftarrow (k + 1) \bmod n$ 
4:   if  $\text{sign}(y_i) \neq \text{sign}(\mathbf{w}^T \mathbf{x}_i)$  then
5:      $\mathbf{w} \leftarrow \mathbf{w} + y_i \mathbf{x}_i$ 
6:   end if
7: until todos los puntos bien clasificados
```

Modificar este pseudo-código para adaptarlo a un algoritmo simple de SVM, considerando que en cada iteración adaptamos los pesos de acuerdo al caso peor clasificado de toda la muestra. Justificar adecuadamente/matematicamente el resultado, mostrando que al final del entrenamiento solo estaremos adaptando los vectores soporte.

4. Considerar un modelo SVM y los siguientes datos de entrenamiento: Clase-1: $\{(1,1), (2,2), (2,0)\}$, Clase-2: $\{(0,0), (1,0), (0,1)\}$
 - a) Dibujar los puntos y construir por inspección el vector de pesos para el hiperplano óptimo y el margen óptimo.
 - b) ¿Cuáles son los vectores soporte?
 - c) Construir la solución en el espacio dual. Comparar la solución con la del apartado (a)
5. Una empresa esta valorando cambiar su sistema de proceso de datos, para ello dispone de dos opciones, la primera es adquirir dos nuevos sistemas idénticos al actual a 200.000 euros cada uno, y la segunda consiste en adquirir un sistema integrado por 800.000 euros.

Las ventas que la empresa estima que tendrá a lo largo de la vida útil de cualquiera de sus equipos son de 5.000.000 de euros en el caso de positivo, a lo que la empresa le asigna una probabilidad de que suceda del 30 %, en caso contrario, las ventas esperadas son de 3.500.000 euros. ¿Que opción debería de tomar la empresa?

6. El método de Boosting representa una forma alternativa en la búsqueda del mejor clasificador respecto del enfoque tradicional implementado por los algoritmos PLA, SVM, NN, etc. a) Identifique de forma clara y concisa las novedades del enfoque; b) Diga las razones profundas por las que la técnica funciona produciendo buenos ajustes (no ponga el algoritmo); c) Identifique sus principales debilidades; d) ¿Cuál es su capacidad de generalización comparado con SVM?
7. ¿Cuál es a su criterio lo que permite a clasificadores como Random Forest basados en un conjunto de clasificadores simples aprender de forma más eficiente? ¿Cuales son las mejoras que introduce frente a los clasificadores simples? ¿Es Random Forest óptimo en algún sentido? Justifique con precisión las contestaciones.