

Trabajo.3: Preguntas de Teoría

Fecha de entrega: 31 Mayo 2017. Valor máximo: 12.5 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. 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. Enuncie las diferencias y semejanzas entre ambas técnicas. Diga cual de ellas considera que nos proporcionará una mejor estimación del error en cada caso concreto y por qué.
2. 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 interando 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.

3. 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)
4. Una empresa esta valorando cambiar su sistema de proceso de datos, para ello dispone de dos opciones, la primera es adquirir un nuevo sistema compuesto por dos sistemas idénticos al actual a 200.000 euros cada uno, y la segunda consiste en adquirir un nuevo sistema mucho mayor por 800.000 euros. Las ventas que la empresa estima que tendrá a lo largo de la vida útil de cualquiera de sus nuevos equipos es de 5.000.000 de euros en el caso de un mercado alcista, 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. Construir el árbol de decisiones y decir que opción es la más ventajosa para la empresa.
5. ¿Que algoritmos de aprendizaje no se afectan por la dimensionalidad del vector de características? Diga cuáles y por qué.
6. Considere la siguiente aproximación al aprendizaje. Mirando los datos, parece que los datos son linealmente separables, por tanto decidimos usar un simple perceptron y obtenemos un error de entrenamiento cero con los pesos óptimos encontrados. Ahora deseamos obtener algunas conclusiones sobre generalización, por tanto miramos el valor d_{vc} de nuestro modelo y vemos que es $d + 1$. Usamos dicho valor de d_{vc} para obtener una cota del error de test.

Argumente a favor o en contra de esta forma de proceder identificando los posible fallos si los hubiera y en su caso cual hubiera sido la forma correcta de actuación.

7. Discuta pros y contras de los clasificadores SVM y Random Forest (RF). Considera que SVM por su construcción a través de un problema de optimización debería ser un mejor clasificador que RF. Justificar las respuestas.
8. ¿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.
9. En un experimento para determinar la distribución del tamaño de los peces en un lago, se decide echar una red para capturar una muestra representativa. Así se hace y se obtiene una muestra suficientemente grande de la que se pueden obtener conclusiones estadísticas sobre los peces del lago. Se obtiene la distribución de peces por tamaño y se entregan las conclusiones. Discuta si las conclusiones obtenidas servirán para el objetivo que se persigue e identifique si hay algo que lo impida.
10. Identifique dos razones de peso por las que el ajuste de un modelo de red neuronal a un conjunto de datos puede fallar o equivalentemente obtener resultados muy pobres. Justifique la importancia de las razones expuestas.