



UNIVERSIDAD DE GRANADA

APRENDIZAJE AUTOMÁTICO
GRADO EN INGENIERÍA INFORMÁTICA

TRABAJO 3

CUESTIONES DE TEORÍA

Autor

Vladislav Nikolov Vasilev

Rama

Computación y Sistemas Inteligentes



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE
TELECOMUNICACIÓN

CURSO 2018-2019

Índice

Ejercicio 1	2
Ejercicio 2	2
Ejercicio 3	2
Ejercicio 4	3
Ejercicio 5	3
Ejercicio 6	3
Ejercicio 7	4
Ejercicio 8	4
Ejercicio 9	4
Ejercicio 10	4
Referencias	6

Ejercicio 1

¿Podría considerarse Bagging como una técnica para estimar el error de predicción de un modelo de aprendizaje? Diga si o no con argumentos. En caso afirmativo compárela con validación cruzada.

Solución

Ejercicio 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 iterando sobre los datos y adaptando los pesos de acuerdo al algoritmo

Algorithm 1 Perceptron

```
1: Entradas:  $(\mathbf{x}_i, y_i) = 1, \dots, n$ ,  $w = 0$ ,  $k = 0$   
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.

Solución

Ejercicio 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 hi-

perplano óptimo y el margen óptimo.

Solución

b) ¿Cuáles son los vectores soporte?

Solución

c) Construir la solución en el espacio dual. Comparar la solución con la del apartado (a)

Solución**Ejercicio 4**

¿Cuál es el criterio de optimalidad en la construcción de un árbol? Analice un clasificador en árbol en términos de sesgo y varianza. ¿Que estrategia de mejora propondría?

Solución**Ejercicio 5**

¿Cómo influye la dimensión del vector de entrada en los modelos: SVM, RF, Boosting and NN?

Solución**Ejercicio 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?

Solución

Ejercicio 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.

Solución

Ejercicio 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.

Solución

Ejercicio 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.

Solución

Ejercicio 10

Identifique que pasos daría y en que orden para conseguir con el menor esfuerzo posible un buen modelo de red neuronal a partir una muestra de datos. Justifique

los pasos propuestos, el orden de los mismos y argumente que son adecuados para conseguir un buen óptimo. Considere que tiene suficientes datos tanto para el ajuste como para el test.

Solución

Referencias

- [1] Texto referencia
<https://url.referencia.com>