

Machine Learning

Tarea #1

Entrega: Septiembre 28 de 2015, 5 pm.

Los datos adjuntos corresponden a un experimento en física en el cual se desea detectar cierto tipo de partículas (clase +1) de la actividad de fondo (clase 0). Se incluyen 15000 datos de entrenamiento de entrada **xtrain** (cada fila corresponde a un dato) y sus correspondientes etiquetas **ytrain**. También se incluyen 5000 datos de entrada de prueba **xtest**. En cada uno de los siguientes casos su objetivo es obtener un modelo que tenga probabilidad de error tan baja como sea posible ¹. Reserve una porción de sus datos (puede ser el 10 %) para estimar el error de clasificación de cada uno de sus modelos finales.

1. Resuelva el problema de clasificación utilizando regresión logística. Estime la probabilidad de error de su modelo. Dé la precisión y confianza de su estimativo.
2. Resuelva el problema de clasificación utilizando un perceptrón multinivel con una capa escondida. Para encontrar el número de neuronas apropiado, usted debe entrenar varias redes con diferente número de neuronas, y elegir la que presente mejor generalización. Para esto debe reservar una parte de sus datos para validar el desempeño de cada una de sus redes². Una posibilidad es seguir el siguiente protocolo:

```
for  $n = 2$  hasta  $N_{MAX}$  do  
    Entrene red con  $n$  neuronas en la capa escondida usando datos de entrenamiento;  
    Evalúe red entrenada en conjunto de validación;  
end  
Escoja red con el menor error en conjunto de validación;
```

Es posible que usted deba experimentar con diferentes funciones de activación, criterios de error y algoritmos de entrenamiento.

Estime la probabilidad de error de su modelo final. Dé la precisión y confianza de su estimativo.

3. Resuelva el problema de clasificación utilizando un perceptrón multinivel con *dos* capas escondidas. Puede emplear un protocolo similar al del punto anterior. Estime la probabilidad de error de su modelo final. Dé la precisión y confianza de su estimativo.

Escriba un informe detallado del procedimiento llevado a cabo para llegar a su modelo final en cada caso, y una comparación de sus resultados en los tres casos. Usted debe adjuntar a su informe las etiquetas que sus modelos finales obtienen en los datos de prueba **xtest**. Con estas etiquetas yo estimaré la probabilidad de error de su modelo. Es válido utilizar el preprocesamiento adicional que le parezca apropiado.

¹Aplican los supuestos usuales de aprendizaje supervisado.

²Estos datos no deben usarse en el entrenamiento, y son diferentes de los datos de prueba que se usan para estimar el error de clasificación final.