

Mentoría Data Science aplicado a BCI

Consignas TP4: Introducción al Aprendizaje Automático

En este Trabajo Práctico utilizaremos la totalidad de las señales del dataset y con todos sus canales para resolver los siguientes ejercicios, procesadas en features y labels según los métodos que plantearon e investigaron en el TP3.

Aprendizaje Automático

Retomando algunos de los puntos del TP anterior que no llegamos a completar, resuelva las siguientes consignas referidas al Aprendizaje Automático de manera generalizada. Estas preguntas son más bien conceptuales, pero sus respuestas pueden variar el desarrollo de las consignas del punto que sigue, por lo que deben analizarlas cuidadosamente:

- A) Entendimiento del problema.
 - a) Está claro que hasta aquí hablamos de un problema de clasificación supervisada, pero ¿a qué subclase dentro de ellos pertenece? (multilabel? multiclase?)
 - b) En palabras, describa cómo podría plantearse un problema de regresión con los datos estudiados.
 - c) De la misma forma, ¿cómo podría pensarse un problema de clasificación no supervisada a partir de los datos disponibles?
- B) En el problema de clasificación supervisada:
 - a) ¿Qué métricas considera que son apropiadas para evaluar el desempeño de algún algoritmo de clasificación? Recuerde limitar la cantidad que elija a un valor pequeño ya que serán usadas para evaluar los algoritmos de la próxima sección.
 - b) ¿Qué funciones de costo consideran apropiadas para entrenar un algoritmo para este problema?

Aprendizaje Supervisado

Ahora sí ingresamos concretamente en lo que respecta a este TP.

Para el equipo 2: Retomen la búsqueda de features informativos, traten de encontrar mejores features y quedense con la técnica que mejor les resulte. Los que presentaron en el TP anterior está claro que no estaban buenos.

Para el equipo 1: Aún a la espera de su TP3.

Resuelva las siguientes consignas.

- A) Benchmarking y desarrollo del algoritmo evaluador:
 - a) Utilice una clasificación aleatoria de los ejemplos para utilizar como benchmark de los resultados posteriores. Este benchmark representa el peor de los desempeños de clasificación posibles.
 - b) Evalúe el desempeño/rendimiento de este benchmark bajo las métricas seleccionadas en el apartado anterior. Considere repetir este procesamiento

algunas veces para obtener un promedio, máximo, mínimo u otro representante de estos resultados, ya que se trata de un proceso completamente aleatorizado.

B) Búsqueda a grandes rasgos:

- a) Utilicen la mayor cantidad de algoritmos de aprendizaje automático supervisado que puedan (mínimo 3). Creen para cada uno un modelo bajo el paradigma de dicho método y entrénelo con el dataset elegido. Opcional: generar curvas de progreso de métricas y funciones de pérdida a lo largo del entrenamiento.
- b) No hace falta que la búsqueda de hiperparámetros extremadamente detallada, ya que esto se llevará a cabo en el inciso C), sino más bien una exploración general de las tendencias de cada modelo.
- c) Si encuentra resultados de las métricas analizadas o gráficos de entrenamiento plotados que resulten destacables -no sólo porque sean valores más óptimos, sino también posibles casos extraños, situaciones de over y underfitting, etc- muéstrellos y analícelos en este inciso.

C) Búsqueda puntualizada:

- a) Con el modelo que presente mejores resultados, lleve a cabo una búsqueda ahora sí más detallista, variando los hiperparámetros y funciones de costo. Si estos métodos permiten variar la cantidad de instancias/épocas de entrenamiento, analice lo que sucede cuando varía las duraciones de entrenamiento.
- b) Reporte los resultados obtenidos y seleccione el set completo de configuraciones que mejor resuelven, bajo su criterio, nuestro problema de clasificación