

Tarea 4 - Análitica con Big Data

Instrucciones

- La pregunta 1 es obligatoria y deberá ser resuelta en grupos.
 - La pregunta 2 es opcional y solamente será desarrollada si se desea reemplazar la nota de alguna otra tarea. Esta pregunta se deberá realizar de forma personal.
 - Se deberá de exponer los resultados en la última semana de clase.
 - La fecha de presentación de los dos trabajos será el día miércoles 20 de Noviembre (en clase). La pregunta 2 (opcional) tendrá como fecha límite el 1 de Diciembre (en clase).
-

Pregunta 1 (20 puntos)

En un trabajo anterior, se le dejó crear un modelo de clasificación de jugadores de fútbol para poder predecir cuál debía ser la mejor posición que este debería de jugar.

Como cualquier modelo de predicción, este está sujeto a errores de generalización (ya sea *overfitting* o *underfitting*), así como a errores por el *curse of dimensionality*. Por este motivo, se le encomienda hacer un análisis detallado de los features a utilizar con el afán de mejorar las predicciones.

Debe exponer sus resultados en un ppt así como entregar su código fuente por blackboard.

Criterios de calificación

Criterio	Puntaje
Elección de features correcto mediante genético	10 puntos
Análisis de resultados	8 puntos
Código (comentarios)	2 puntos

Pregunta 2 (20 puntos, Opcional)

Los algoritmos genéticos nos permiten encontrar soluciones útiles a problemas donde el espacio de búsqueda es bastante grande.

Para la técnicas de machine learning, especialmente en aprendizaje supervisado, se utilizan algoritmos que nos permiten encontrar una solución óptima global (algoritmos de optimización). Uno de estos algoritmos es el llamado Gradient Descent.

En clase, se estudiaron los algoritmos genéticos que también nos dan una manera de encontrar una solución óptima global para un espacio de búsqueda grande.

Se le pide implementar un programa en python que nos permita encontrar el modelo h (parámetros Theta) que minimicen el error de la función predictora.

Puede utilizar las siguientes referencias:

- https://www.researchgate.net/publication/235949653_Genetic_algorithm_optimization_technique_for_linear_regression_model
- <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.333.1528&rep=rep1&type=pdf>