



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA
DEPARTAMENTO DE CIENCIA DE LA COMPUTACIÓN

IIC2233 - Programación Avanzada
2° semestre 2015

Actividad 28

Visualización y análisis de datos

Instrucciones

Se le ha confiado una base de datos confidencial (`exámenes.csv`) que contiene los resultados de los exámenes de muchos pacientes del Hospital CCD. Esta base de datos además cuenta con una columna con el diagnóstico del paciente (sano o enfermo). En esta actividad usted deberá programar un algoritmo que actúe como juez (médico diagnosticador), es decir, que tome el resultado de los exámenes y diga sin intervención humana si el paciente está sano o enfermo.

Requerimientos

Se debe importar con la librería **pandas** la base de datos **exámenes.csv** (ver `pandas.read_csv()`). Para ver su contenido imprima las columnas del `dataFrame` generado.

No debe subir ninguno de los archivos CSV entregados, de subirlos será penalizado.

To - DO

- (1.50 pts) Grafique en dos dimensiones todos los pares de exámenes de la base de datos (`examen0` vs `examen1`, `examen0` vs `examen2`, etc..). Graficando con diferente color los datos de los pacientes sanos y los de los enfermos (ejemplo verde-rojo).
- (1.50 pts) Ajuste un polinomio de grado 1 o 2 (ver `numpy.polyfit()`) en cada plano encontrado en el paso anterior y grafíquelo en el plano (nótese que si los datos son muy dispersos este ajuste no le hará mucho sentido).
- (1.50 pts) Elija el mejor espacio encontrado en el que se permita diferenciar claramente sanos de enfermos. Y proponga un método de clasificación entre sano y enfermo si se conocen todos los exámenes de un paciente.
- (1.50 pts) Para cada paciente en **pacientes_nuevos.csv** imprima su clasificación y grafique a este paciente en el espacio elegido anteriormente, con un color que permita distinguirlo (ejemplo negro) - cree un gráfico por cada paciente.