

Obligatorio 4 - Aprendizaje Estadístico

Para la planificación de una red es de suma importancia conocer (o al menos estimar) la demanda de tráfico que será inyectada a la misma. Es decir, cuánto tráfico por unidad de tiempo se pretende traficar. Una buena planificación de la misma no sólo considera el valor actual de la demanda, sino que también toma en cuenta posibles valores futuros para evitar realizar cambios en el despliegue en el futuro cercano.

En este obligatorio nos concentraremos en esta última etapa: la de proyección a futuro. Consideraremos que ya fue realizada la (no sencilla) tarea de recolectar datos históricos sobre la demanda y se pide dar una estimación a futuro de cuál será el valor de la misma. En particular, se tienen medidas de la demanda media cada cinco minutos durante 4 semanas (ver archivo traza.rar en la página del curso). Se pide dar una predicción de la demanda para la semana siguiente, i.e. un valor cada cinco minutos durante los siguientes 7 días.

Tal como sucedió en obligatorios anteriores se puede utilizar cualquier herramienta disponible para realizar la predicción. A diferencia de obligatorios anteriores, esta vez sí existe una respuesta correcta (aunque no acerca del mejor método para predecirla). Por ende, se plantea una (esperemos que sana) competencia entre los estudiantes. Las entregas serán puntuadas según el error absoluto medio \bar{E} definido de la siguiente forma:

$$\bar{E} = \frac{1}{N} \sum_i^N |X_i - \hat{X}_i|$$

donde X_i es el verdadero valor de la demanda y \hat{X}_i su predicción correspondiente.

Aquel estudiante cuyo método obtenga el menor \bar{E} recibirá una nota de 100. Aquel cuyo método obtenga el mayor \bar{E} recibirá un 70. Los estudiantes cuyo \bar{E} esté entre ambos recibirán una nota entre 70 y 100 lineal con respecto al error obtenido. Esto no significa que el 70 esté asegurado, pues se deberá entregar un informe que describa la metodología que los llevó a obtener la predicción entre otras cosas.

En detalle se pide que se entregue en un único archivo lo siguiente:

- El script con el cual se puede obtener la predicción. Éste deberá tomar como entrada el archivo descomprimido mencionado antes y devolver un archivo como se describe en el siguiente punto. Recordar que no hay porqué limitarse a MATLAB y que se puede utilizar cualquier otro programa (incluir un pequeño instructivo de cómo se usa el script). En la página del curso se dejan enlaces y ejemplos de otras posibles herramientas.
- Un archivo con los valores de la predicción cada cinco minutos. Este archivo debe contener en cada línea el valor de la predicción únicamente (de forma tal que sea fácilmente parseable por algún programa tipo MATLAB). Este archivo se entregará como respaldo. Siempre que sea posible los docentes reproducirán la predicción con el script del punto anterior.

- Un informe más o menos detallado de la metodología utilizada para realizar la predicción (no más de 10 carillas, idealmente no más de 5). Éste debe incluir la descripción de al menos: pre-procesamiento realizado a la traza (si hubo alguno), justificación de la elección del algoritmo de predicción incluyendo su parametrización, análisis de los resultados obtenidos y justificación de porqué cree que ganará la competencia.

¡Que gane el mejor!