



Universidad Autónoma de Baja California

Maestría y Doctorado en Ciencias e Ingeniería
Ingeniería en Computación

6. Regresión

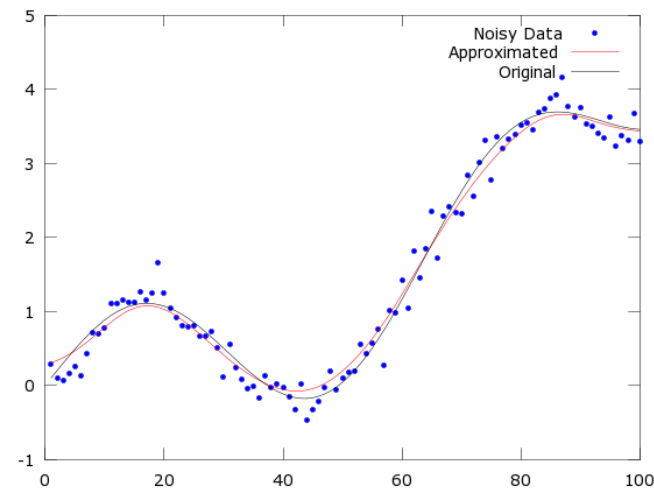
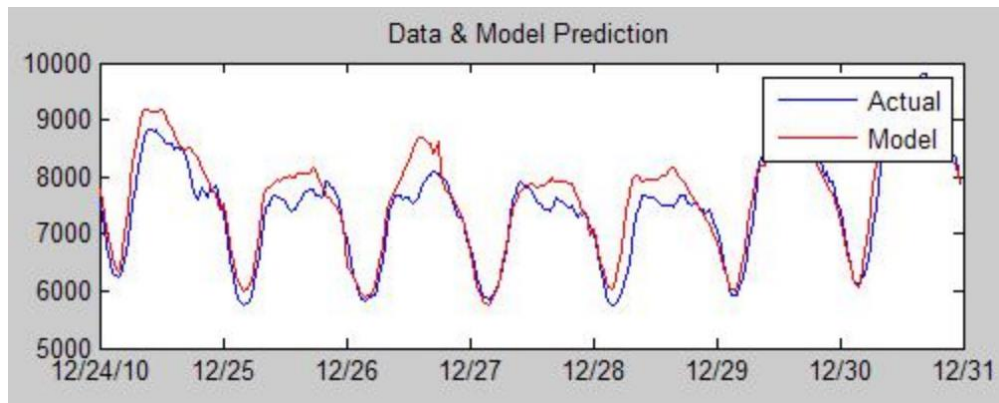
Minería de Datos

¿Qué es regresión?

- ▶ Es una forma de análisis de datos que extrae modelos que describen comportamientos a partir de datos. Entiéndase por "comportamientos" a una serie de números reales que siguen cierto patrón a partir de sus atributos.

Conceptos básicos de regresión

- ▶ Ejemplos de modelos de regresión y sus usos...
- ▶ ¿Cómo funciona la regresión? Se realiza en dos pasos, que consisten en un paso de entrenamiento (donde se construye el modelo de regresión) y uno de identificación (donde el modelo es utilizado para predecir comportamientos para un conjunto dado de datos).
- ▶ Es análogo a clasificación, pero en lugar de identificar clases finitas, se identifican comportamientos continuos.



Índices de validación

- ▶ R (linear regression)
- ▶ R^2 (coefficient of determination)
- ▶ SSE (sum of squared errors)
- ▶ MAE (mean absolute error)
- ▶ MSE (mean squared error)
- ▶ RMSE (root mean squared error)

Actividad

- ▶ Trabajar con 3 base de datos distintas
 - ▶ Que sean de 1, 2, y +10 salidas.
- ▶ Aplicar todo índice de validación a cada modelo de cada dataset
 - ▶ Cada quién escogerá y explicará un índice diferente para exponer: ¿qué te dice el índice?
- ▶ Separando los datos mediante:
 - ▶ Hold-Out (60/40)
 - ▶ Random subsampling (N=30)
 - ▶ K-fold (K=5 y 10)
- ▶ Utilizar 3 técnicas diferentes para generar los modelos de cada caso.
 - ▶ Donde cada técnica tiene un grado de complejidad diferente (por ej. Regresión lineal/polinomial/logarítmica/etc, árbol de decisión para regresión, red neurona, etc.)
 - ▶ Explicar de manera general cómo funciona cada técnica (ponerse de acuerdo para escoger una cada uno)
- ▶ Identificar el mejor resultado por dataset