

Auxiliar 7 - K-NN y Regresión 2

Cátedra: Inteligencia Artificial

Profesor: Gonzalo Ríos

Auxiliar: Miguel Romero

02 de Junio del 2011

- Este ejercicio es un repaso de algunos conceptos en minería de datos
 - Nombre y explique las tareas en minería de datos vistas en clase, incluyendo sus funciones objetivas.
 - Indique al menos 3 métodos de evaluación, mencionando sus ventajas y desventajas.
 - Explique el funcionamiento del algoritmo k -means para clustering.
 - Explique la importancia de la función de error y del espacio de funciones en un problema de regresión funcional.
 - Cuando es adecuado utilizar funciones discriminantes para clasificar? Explique brevemente como funciona esta técnica.
- Tenemos un conjunto de datos de entrenamiento $\{(\vec{x}_i, y_i) : i = 1, \dots, m\}$, donde $\vec{x}_i \in \mathbb{R}^n$ e $y_i \in \mathbb{R}$.
 - Deduzca que el regresor lineal que minimiza el error cuadrático está dado por $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$, con
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad \hat{\beta}_1 = \frac{\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^m (x_i - \bar{x})^2}$$
donde, $\bar{x} = \frac{1}{m} \sum_{i=1}^m x_i$ e $\bar{y} = \frac{1}{m} \sum_{i=1}^m y_i$, son las medias muestrales.
 - Suponga que tiene dos atributos X_1 y X_2 de una persona, que toman valores en \mathbb{R} , y la variable Y que desea entender es la probabilidad de que la persona sea un asesino. Tenemos un conjunto de datos de entrenamiento $\{(x_i^1, x_i^2, y_i) : i = 1, \dots, m\}$. Cual es el problema con aplicar directamente regresión lineal? Como ocuparía la función $f(z) = \frac{1}{1+e^{-z}}$?
- Queremos predecir el precio del dolar de mañana. Como datos tenemos el precio del dolar de cada día de los últimos 10 años. Si asumimos que el precio de cierto día solo depende de los 3 días anteriores, de una manera lineal, como resolvería este problema?
- Responda las siguientes preguntas
 - Cual es el método básico para hacer regresión utilizando el algoritmo k -NN, si cada vecino pesa lo mismo?.
 - Nuevamente, considerando que cada vecino pesa lo mismo. Como mezclaría el algoritmo k -NN con regresión lineal?
 - Finalmente, si cada vecino tiene un peso según la cercanía a la nueva instancia, Como mezclaría el algoritmo k -NN con regresión lineal?