29/04/2024

Regresión logística

Crea una recta y le aplica la función sigmoide para romper la linealidad y que se ajuste mejor a los resultados. Se utiliza para problemas de clasificación, en lugar de la regresión lineal que es para problemas de regresión

kNN

Clasificación en base al vecino más cercano (Nearest neighbour). La ‘k’ define el número de vecinos más cercanos. Este algoritmo no tiene apenas entrenamiento por lo que es muy rápido. Sus principales desventajas son los posibles empates que pueden ocurrir si encuentra el mismo número de vecinos. Otro problema es saber el número de ‘k’.

Para utilizar este modelo los datos han de estar normalizados o estandarizados

Para saber la k adecuada se puede utilizar la formula k = raíz(n)

Medidas de distancia

Distancia euclídea: Pitágoras

Distancia manhattan: sumando catetos. Útil para problemas de navegación con el coche

Distancia Minkowski: un punto medio entre euclídea y manhattan

06/05/2024

Modelo SVM

Crea una recta que separa las dos clases. Utiliza la técnica del vector soporte para buscar la línea mas ancha entre las dos clases.

Cuando los datos estas perfectamente separados, se utiliza un separador de margen duro. Se los datos están solapados se debe aplicar un clasificador de margen blando

Datos linealmente no separables

Cuando los datos no se pueden separar linearmente, se utiliza el truco kernel. Se agrega un polinomio para que los datos pasen de ser rectos a una curva