26/02/2024

Reducion de dimensionalidad

Existen dos formas de reducer la dimensionalidad:

* Filtro:
* Envoltura: itera para buscar la mejora característica
* Métodos integrados:
* Métodos hibirdos

Métodos de filtro

Existen dos categorías:

* Filtro univariante: no tiene en cuenta la relación entre las características. Busca las características con poca desviación típica. Con la funcion VarianceThreshold, se puede realizar automáticamente
* Filtro multivariante: eliminar todas las características que eliminene cierto umbral de correlación entre si. Hay que tener en cuenta que la reacion también puede ser negativa.

04/03/2024

Reducción de dimensionalidad

Búsqueda voraz

Selección hacia adelante

* P-value: la probabilidad de que la hipótesis nula sea verdadera
* Internamente cuando se hace un algoritmo también se hace el contraste de hipótesis para el peso de la características
* Esta reducción de dimensionalidad entrena un modelo con todas la características individualmente, escoge la que de el p-value menor, y luego vuelve a iterar

Esta técnica tiene varios problemas:

* Computacionalmente es costoso
* Tienes que indicarle tu el numero de características que deseas seleccionar
* Dado su funcionamiento, no siempre elige las mejores características

Eliminacion hacia atrás

Exactamente igual que el descrito anteriormente, pero este, en su lugar, coge todas la características al principio y va eliminando la característica que peor métrica le da.

Selección bidireccional

Una mezcla de los dos anteriores. Va añadiendo características, pero cada vez que añade también considera borrar alguna característica para ver si puede borrar alguna característica previamente seleccionada