

Examen Primer Parcial

Ciencia de Datos y Machine Learning - Lic. Patricia Rodríguez

Nombre Completo *

Pablo Jesus Pardo Alcocer

Parte I. Responder de manera concreta

Cuáles los componentes de un algoritmo de aprendizaje?

Los componentes son: los datos de entrenamiento, los datos de validación y el modelo del algoritmo

Cuál la clasificación de aprendizaje automático?

Se clasifica en 4 categorías: aprendizaje supervisado, aprendizaje semisupervisado, aprendizaje no supervisado y aprendizaje por refuerzo.

Cuales las actividades principales dentro del proceso de aprendizaje automático?

Inicialmente se hace la recopilación de datos, luego se hace la limpieza de estos datos, luego se hace la extracción y selección de elementos, también se debe crear el modelo para hacer el entrenamiento con estos datos, luego se hace la evaluación del modelo (generalmente se evalúa el error y la exactitud) y posteriormente se hace la integración y el despliegue del modelo y por último se realiza una retroalimentación e iteración sobre los pasos mencionados.



Qué es un dataset?

Un dataset es una colección de datos que contiene varios registros, donde cada uno de estos registros se lo denomina como una muestra y los atributos particulares en cada registro se llaman características. Generalmente de este dataset es de donde se extraen 2 conjuntos importantes para el entrenamiento y la validación del modelo, que son el set de entrenamiento y el set de pruebas respectivamente.

Cuáles las actividades principales dentro el preprocesamiento de datos?

La limpieza y normalización de las columnas, la limpieza y normalización de las filas y el filtrado de columnas

Cuáles las razones por las que es necesario el seleccionar las características de los datos?

De esta manera se tiene un modelo mucho más optimizado, también el modelo durante el entrenamiento podría tomar en cuenta datos irrelevantes que podrían afectar en su desempeño si es que esta selección de características no se la hace de manera correcta, esto podría conducir a un subajuste del modelo.

Cuáles los métodos para seleccionar las características de los datos?

Existen muchas técnicas, entre ellas la eliminación de filas, la eliminación de columnas, también en caso de que a algún registro le falte alguna característica se puede utilizar la media de todos los demás registros, sin embargo esta última técnica lleva algo de riesgo.

Cuáles las características a evaluar en la selección de un buen modelo?

Generalmente se evalúa sobre todo la exactitud, la tasa de reconocimiento, la tasa positiva real, la especificidad y tasa negativa real, la sensibilidad a cambios y la precisión del modelo



Cuál la función de pérdida de la regresión lineal ?

La función de pérdida de la regresión lineal es: $J(x) = (1/(2m)) * \text{Sumatoria}((h_w(x) - y)^2)$

A qué problemas se aplica la función softmax?

Se aplica a problemas de clasificación multi-clase, o sea a problemas donde se tenga más de dos clases para su clasificación por el modelo ya que la función Softmax evalúa todas las probabilidades y devuelve la clase que tenga la probabilidad más alta.

Parte II. Desarrolle según el enunciado

Usando el dataset adjunto, realizar los siguientes procesos y mostrar los resultados: a) Limpieza de datos, especificar cuales las actividades y adjuntar el dataset de entrenamiento resultante; b) Aplicar un modelo regresional el mas adecuado y entrenar el modelo; c) Validar el modelo obtener los parámetros de exactitud, sensibilidad y precisión



diabetesIncisoA ...

Este formulario se creó en Universidad Mayor de San Simón.

Google Formulario

