**T81-558: Aplicaciones de redes neuronales profundas**

**Módulo 14: Otras técnicas de redes neuronales**

* Instructor: Jeff Heaton, Escuela de Ingeniería McKelvey, Universidad de Washington en St. Louis
* Para obtener más información, visite el sitio web de la clase.

Material de vídeo del módulo 14

* Parte 14.1: Qué es AutoML [Video] [Notebook]
* Parte 14.2: Uso de codificadores automáticos de eliminación de ruido en Keras [Video] [Notebook]
* Parte 14.3: Entrenamiento de un sistema de detección de intrusiones con KDD99 [Video] [Notebook]
* Parte 14.4: Detección de anomalías en Keras [Video] [Notebook]
* Parte 14.5: Las tecnologías de aprendizaje profundo que me entusiasman [Video] [Notebook]

**Parte 14.4: Entrenamiento de un sistema de detección de intrusos con KDD99**

El conjunto de datos KDD-99 es muy famoso en el campo de la seguridad y casi un "hola mundo" de los sistemas de detección de intrusiones (IDS) en el aprendizaje automático. Un sistema de detección de intrusiones (IDS) es un programa que monitorea las computadoras y los sistemas de red para detectar actividades maliciosas o violaciones de políticas. Por lo general, cualquier actividad de intrusión o infracción se informa a un administrador o se recopila de forma centralizada. Los tipos de IDS varían en alcance desde computadoras individuales hasta grandes redes. Aunque el conjunto de datos KDD99 tiene más de 20 años, todavía se usa ampliamente para demostrar los sistemas de detección de intrusiones (IDS). KDD99 es el conjunto de datos utilizado para el Tercer Concurso Internacional de Herramientas de Minería de Datos y Descubrimiento de Conocimiento, que se llevó a cabo en conjunto con KDD-99 La Quinta Conferencia Internacional sobre Descubrimiento de Conocimiento y Minería de Datos. La tarea de la competencia fue construir un detector de intrusiones en la red, un modelo predictivo capaz de distinguir entre conexiones "malas", llamadas intrusiones o ataques, y conexiones normales "buenas". Esta base de datos contiene un conjunto estándar de datos para ser auditados, incluida una amplia variedad de intrusiones simuladas en un entorno de red militar.

Leer en conjunto de datos KDD-99 sin procesar

El siguiente código lee el conjunto de datos CSV KDD99 en un marco de datos de Pandas. El formato estándar de KDD99 no incluye nombres de columna. Por eso, el programa los agrega.

**Analizar un conjunto de datos**

Antes de preprocesar el conjunto de datos KDD99, echemos un vistazo a las columnas y distribuciones individuales. Puede usar la siguiente secuencia de comandos para brindar una descripción general de alto nivel de cómo aparece un conjunto de datos.

import pandas as pd

import os

import numpy as np

from sklearn import metrics

from scipy.stats import zscore

El análisis analiza cuántos valores únicos están presentes. Por ejemplo, la duración, que es un valor numérico, tiene 2495 valores únicos y hay una superposición del 0%. Un valor text / categórico como protocol\_type solo tiene unos pocos valores únicos, y el programa muestra los porcentajes de cada uno. Las columnas con una gran cantidad de valores únicos no muestran sus recuentos de elementos para ahorrar espacio de visualización.

# Analyze KDD-99

analyze(df)

# Codificar el vector de características[¶](http://localhost:8888/notebooks/Desktop/Proyecto_IDS/Untitled%20Folder/IDS_2.ipynb#Codificar-el-vector-de-caracter%C3%ADsticas)

Usamos las mismas dos funciones proporcionadas anteriormente para preprocesar los datos. El primero codifica Z-Scores y el segundo crea variables ficticias a partir de columnas categóricas.

Intentaremos predecir qué tipo de ataque se está llevando a cabo. La columna de resultados especifica el tipo de ataque. Un valor de normal indica que no hay ningún ataque en curso. Mostramos los resultados; algunos tipos de ataques son mucho más raros que otros.

# Entrene la red neuronal

Ahora entrenamos la red neuronal para clasificar los diferentes resultados de KDD99. El código proporcionado aquí implementa un sistema neuronal relativamente simple con dos capas ocultas. Lo entrenamos con los datos KDD99 proporcionados.