DADOS DO PROJETO 03 - CLASSIFICAÇAO DE CURVAS DE LUZ

Os dados estao em plasticc\_train\_(lightcurves, metadata) e plasticc\_lightcurves\_extra, plasticc\_test\_metadata. note1\_dataRelease contem informacoes sobre os dados; the-plasticc-astronomy-starter-kit apresenta o contexto do challenge, um pouco de astronomia, e tem algumas funçoes para ajudar a visualizar as curvasde luz.

Inicialmente, os dados a serem utilizados terao a tag "train", e os com a tag "teste" e "extra" sao para testar os resultados apenas.

Os arquivos "lightcurve" contem a curva de luz em diferentes filtros (ugrizY). Cada linha tem um "object\_id" correspondente em "metadata", que pode ser utilizada para fazer o match com a respectiva classe ("target"), e outros metadados que porventura possam ser uteis. Cada id e unica e serve para identificaçao dos objetos, portanto a curva de luz que um determinado objeto corresponde a todas as linhas em plasticc\_train\_lightcurves com a id correspondete ao objeto, Lembrando que qualquer coluna dos metadados que tenha "true" no nome nao pode ser usado para a classificaçao.

Para facilitar e ajudar na compreensao, segue um dicionario mapeando a classe numerica ao nome dos objetos astronomicos

{

6: 'Single micro-lens - 6',

15: 'TDE - 15',

16: 'Eclipsing Binary - 16',

42: 'SNII - 42',

52: 'SNIax - 52',

53: 'Mira - 53',

62: 'SNIbc - 62',

64: 'Kilonova - 64',

65: 'M-dwarf - 65',

67: 'SNIa-91bg - 67',

88: 'AGN - 88',

90: 'SNIa - 90',

92: 'RR lyrae - 92',

95: 'SLSN-I - 95',

99: 'Unknown - 99',

}

Para facilitar, algumas classes podem ser agrupadas em super classes. A seguir, podem usar o dicionário que mapeia das classes menos abrangentes para as classes mais abrangentes

{42: 0,

52: 0,

62: 0,

67: 0,

90: 0,

6: 1,

64: 1,

65: 1,

15: 2,

95: 2,

16: 3,

53: 3,

92: 3,

88:4

}

Nomes das classes abrangentes:

{0: ‘SN-like’,

1: ‘Fast’,

2: ‘Long’

3: ‘Periodic’,

4: ‘Non-periodic’

}

SN-like - tipos de supernova

Fast - curva de luz sobe e desce rápido

Long - curva de luz sobe a desce mais devagar

Periodic - curva de luz sobe e desce periodicamente

Non-Periodic - curva de luz aparece e some sem previsão

A classe 99 so esta presente no set de teste, e pode ser ignorada.

O sample de teste e um corte no sample total, com 20.000 objetos; portanto, nem todo object\_id nos metadados tera um correspondente em lightcurves\_extra

DICAS:

- Usem o pandas para abrir os arquivos .csv.gz sem precisar descomprimir (menos memoria usada) e converter para numpy/tf arrays na hora de montar as curvas de luz

- Fazer uma analise rapida antes de começar para ver algumas coisas, como: distribuiçao de classes, tamanho das curvas de luz, plotar alguns exemplos, ...

- A informaçao sobre os filtros e PROVAVELMENTE importante! Nao perde-la montando os dados para entrar na rede

Possível melhoras:

TimeDistributed(Dense(...)) antes da BiLSTM