

```

!/usr/bin/env python3
!pip install -U deepctr-torch
import pandas as pd
import torch
from sklearn.metrics import log_loss, roc_auc_score
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, MinMaxScaler

from deepctr_torch.inputs import SparseFeat, DenseFeat, get_feature_names
from deepctr_torch.models import *

from google.colab import drive
import matplotlib.pyplot as plt

```

```

Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch)
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch)
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch)
Requirement already satisfied: nvidia-cudnn-cu12==8.9.2.26 in /usr/local/lib/python3.10/dist-packages (from torch>=1.2)
Requirement already satisfied: nvidia-cublas-cu12==12.1.3.1 in /usr/local/lib/python3.10/dist-packages (from nvidia-cusolver==1.
Requirement already satisfied: nvidia-cufft-cu12==11.0.2.54 in /usr/local/lib/python3.10/dist-packages (from torch==1.
Requirement already satisfied: nvidia-curand-cu12==10.3.2.106 in /usr/local/lib/python3.10/dist-packages (from torch>=
Requirement already satisfied: nvidia-cusolver-cu12==11.4.5.107 in /usr/local/lib/python3.10/dist-packages (from torch
Requirement already satisfied: nvidia-cuspars-cu12==12.1.0.106 in /usr/local/lib/python3.10/dist-packages (from torch
Requirement already satisfied: nvidia-nccl-cu12==2.19.3 in /usr/local/lib/python3.10/dist-packages (from torch==1.2.0-
Requirement already satisfied: nvidia-nvtx-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch==1.2.
Requirement already satisfied: triton==2.2.0 in /usr/local/lib/python3.10/dist-packages (from torch==1.2.0->deepctr-to
Requirement already satisfied: nvidia-nvjitlink-cu12 in /usr/local/lib/python3.10/dist-packages (from nvidia-cusolver-
Requirement already satisfied: numpy>=1.17.3 in /usr/local/lib/python3.10/dist-packages (from scikit-learn->deepctr-to
Requirement already satisfied: scipy>=1.3.2 in /usr/local/lib/python3.10/dist-packages (from scikit-learn->deepctr-to
Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from scikit-learn->deepctr-to
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn->dee
Requirement already satisfied: absl-py>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow->deepctr-tor
Requirement already satisfied: astunparse>=1.6.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow->deepctr-
Requirement already satisfied: flatbuffers>=23.5.26 in /usr/local/lib/python3.10/dist-packages (from tensorflow->deepc
Requirement already satisfied: gast!=0.5.0,!>=0.5.1,!>=0.5.2,>=0.2.1 in /usr/local/lib/python3.10/dist-packages (from te
Requirement already satisfied: google-pasta>=0.1.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow->deepctr
Requirement already satisfied: h5py>=2.9.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow->deepctr-torch)
Requirement already satisfied: libclang>=13.0.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow->deepctr-t
Requirement already satisfied: ml-dtypes<=0.2.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow->deepctr-t
Requirement already satisfied: opt-einsum>=2.3.2 in /usr/local/lib/python3.10/dist-packages (from tensorflow->deepctr-
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from tensorflow->deepctr-torch) (
Requirement already satisfied: protobuf!=4.21.0,!>=4.21.1,!>=4.21.2,!>=4.21.3,!>=4.21.4,!>=4.21.5,<5.0.0dev,>=3.20.3 in /u
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from tensorflow->deepctr-torch)
Requirement already satisfied: six>=1.12.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow->deepctr-torch)
Requirement already satisfied: termcolor>=1.1.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow->deepctr-t
Requirement already satisfied: wrapt<1.15,>=1.11.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow->deepctr-t
Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.23.1 in /usr/local/lib/python3.10/dist-packages (from t
Requirement already satisfied: grpcio<2.0,>=1.24.3 in /usr/local/lib/python3.10/dist-packages (from tensorflow->deepctr
Requirement already satisfied: tensorboard<2.16,>=2.15 in /usr/local/lib/python3.10/dist-packages (from tensorflow->de
Requirement already satisfied: tensorflow-estimator<2.16,>=2.15.0 in /usr/local/lib/python3.10/dist-packages (from ten
Requirement already satisfied: keras<2.16,>=2.15.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow->deepctr
Requirement already satisfied: wheel<1.0,>=0.23.0 in /usr/local/lib/python3.10/dist-packages (from astunparse>=1.6.0->
Requirement already satisfied: google-auth<3,>=1.6.3 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.16
Requirement already satisfied: google-auth-oauthlib<2,>=0.5 in /usr/local/lib/python3.10/dist-packages (from tensorboa
Requirement already satisfied: markdown>=2.6.8 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.16,>=2.1
Requirement already satisfied: requests<3,>=2.21.0 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.16,>
Requirement already satisfied: tensorboard-data-server<0.8.0,>=0.7.0 in /usr/local/lib/python3.10/dist-packages (from
Requirement already satisfied: werkzeug>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.16,>=2.1
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from jinja2->torch==1.2.0->
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/dist-packages (from sympy->torch==1.2.0->deep
Requirement already satisfied: cachetools<6.0,>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>
Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>
Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>=1.6.3->t
Requirement already satisfied: requests-oauthlib>=0.7.0 in /usr/local/lib/python3.10/dist-packages (from google-auth-o
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tens
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0
Requirement already satisfied: pyasn1<0.7.0,>=0.4.6 in /usr/local/lib/python3.10/dist-packages (from pyasn1-modules>=0
Requirement already satisfied: oauthlib>=3.0.0 in /usr/local/lib/python3.10/dist-packages (from requests-oauthlib>=0.7

```

```
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=
```

```
data = pd.read_csv(r'drive/My Drive/filtered_train.csv')
data = data.iloc[:, 1:]
```

```
data.shape
```

```
(404290, 24)
```

```
data.dtypes
```

```
id                float64
click             int64
hour             int64
C1               int64
banner_pos       int64
site_id          object
site_domain      object
site_category    object
app_id           object
app_domain       object
app_category     object
device_id        object
device_ip        object
device_model     object
device_type      int64
device_conn_type int64
C14             int64
C15             int64
C16             int64
C17             int64
C18             int64
C19             int64
C20             int64
C21             int64
dtype: object
```

```
data.nunique()
```

```
id                402586
click             2
hour             240
C1               7
banner_pos       7
site_id          2225
site_domain      2188
site_category    22
app_id           2241
app_domain       143
app_category     27
device_id        64742
device_ip        261706
device_model     4380
device_type      4
device_conn_type 4
C14             2088
C15             8
C16             9
C17             411
C18             4
C19             65
C20             161
C21             60
dtype: int64
```

```
data['click'].value_counts()
```

```
click
0    335635
1    68655
Name: count, dtype: int64
```

```

sparse_features = list(data.select_dtypes(['object']).columns)
dense_features = list(data.select_dtypes(['int64']).columns)
dense_features.remove('click')
sparse_features.remove('device_id')
sparse_features.remove('device_ip')

data[sparse_features] = data[sparse_features].fillna('-1', )
data[dense_features] = data[dense_features].fillna(0, )
target = ['click']

for feat in sparse_features:
    lbe = LabelEncoder()
    data[feat] = lbe.fit_transform(data[feat])
mms = MinMaxScaler(feature_range=(0, 1))
data[dense_features] = mms.fit_transform(data[dense_features])

fixlen_feature_columns = [SparseFeat(feat, data[feat].nunique())
                           for feat in sparse_features] + [DenseFeat(feat, 1, )
                                                             for feat in dense_features]

print(fixlen_feature_columns)
dnn_feature_columns = fixlen_feature_columns
linear_feature_columns = fixlen_feature_columns

fixlen_feature_columns = [SparseFeat(feat, data[feat].nunique())
                           for feat in sparse_features] + [DenseFeat(feat, 1, )
                                                             for feat in dense_features]
dnn_feature_columns = [DenseFeat(feat, 1, ) for feat in dense_features]
linear_feature_columns = [SparseFeat(feat, data[feat].nunique()) for feat in sparse_features]

feature_names = get_feature_names(linear_feature_columns + dnn_feature_columns)

[SparseFeat(name='site_id', vocabulary_size=2225, embedding_dim=4, use_hash=False, dtype='int32', embedding_name='site_id')
]

train, test = train_test_split(data, test_size=0.2, random_state=2022)
train_model_input = {name: train[name] for name in feature_names}
test_model_input = {name: test[name] for name in feature_names}

# print(train_model_input[0,:])
# print(test[target].value_counts())

device = 'cpu'
use_cuda = True
if use_cuda and torch.cuda.is_available():
    print('cuda ready...')
    device = 'cuda:0'

model = DeepFM(linear_feature_columns=linear_feature_columns, dnn_feature_columns=dnn_feature_columns,
               task='binary',
               l2_reg_embedding=1e-5, device=device)

model.compile("adagrad", "binary_crossentropy",
             metrics=["binary_crossentropy", "auc"], )

history = model.fit(train_model_input, train[target].values, epochs=30, verbose=2, validation_split=0.2)
pred_ans = model.predict(test_model_input, 256)
print("")
print("test LogLoss", round(log_loss(test[target].values, pred_ans), 4))
print("test AUC", round(roc_auc_score(test[target].values, pred_ans), 4))

cpu
Train on 258745 samples, validate on 64687 samples, 1011 steps per epoch
Epoch 1/30
22s - loss: 0.4277 - binary_crossentropy: 0.4277 - auc: 0.6856 - val_binary_crossentropy: 0.4207 - val_auc: 0.698
Epoch 2/30
13s - loss: 0.4178 - binary_crossentropy: 0.4178 - auc: 0.7111 - val_binary_crossentropy: 0.4166 - val_auc: 0.707
Epoch 3/30
12s - loss: 0.4146 - binary_crossentropy: 0.4146 - auc: 0.7178 - val_binary_crossentropy: 0.4145 - val_auc: 0.711
Epoch 4/30
12s - loss: 0.4127 - binary_crossentropy: 0.4126 - auc: 0.7210 - val_binary_crossentropy: 0.4132 - val_auc: 0.713
Epoch 5/30
14s - loss: 0.4113 - binary_crossentropy: 0.4113 - auc: 0.7240 - val_binary_crossentropy: 0.4123 - val_auc: 0.715
Epoch 6/30
13s - loss: 0.4104 - binary_crossentropy: 0.4104 - auc: 0.7255 - val_binary_crossentropy: 0.4116 - val_auc: 0.716
Epoch 7/30

```

```

14s - loss: 0.4096 - binary_crossentropy: 0.4096 - auc: 0.7274 - val_binary_crossentropy: 0.4110 - val_auc: 0.717
Epoch 8/30
13s - loss: 0.4089 - binary_crossentropy: 0.4089 - auc: 0.7285 - val_binary_crossentropy: 0.4105 - val_auc: 0.718
Epoch 9/30
14s - loss: 0.4083 - binary_crossentropy: 0.4083 - auc: 0.7294 - val_binary_crossentropy: 0.4101 - val_auc: 0.719
Epoch 10/30
14s - loss: 0.4078 - binary_crossentropy: 0.4078 - auc: 0.7308 - val_binary_crossentropy: 0.4098 - val_auc: 0.719
Epoch 11/30
12s - loss: 0.4074 - binary_crossentropy: 0.4074 - auc: 0.7314 - val_binary_crossentropy: 0.4095 - val_auc: 0.720
Epoch 12/30
13s - loss: 0.4070 - binary_crossentropy: 0.4070 - auc: 0.7318 - val_binary_crossentropy: 0.4093 - val_auc: 0.720
Epoch 13/30
13s - loss: 0.4066 - binary_crossentropy: 0.4066 - auc: 0.7325 - val_binary_crossentropy: 0.4090 - val_auc: 0.721
Epoch 14/30
14s - loss: 0.4063 - binary_crossentropy: 0.4063 - auc: 0.7331 - val_binary_crossentropy: 0.4087 - val_auc: 0.721
Epoch 15/30
13s - loss: 0.4060 - binary_crossentropy: 0.4060 - auc: 0.7339 - val_binary_crossentropy: 0.4085 - val_auc: 0.722
Epoch 16/30
15s - loss: 0.4057 - binary_crossentropy: 0.4057 - auc: 0.7341 - val_binary_crossentropy: 0.4083 - val_auc: 0.722
Epoch 17/30
14s - loss: 0.4054 - binary_crossentropy: 0.4054 - auc: 0.7346 - val_binary_crossentropy: 0.4082 - val_auc: 0.722
Epoch 18/30
13s - loss: 0.4052 - binary_crossentropy: 0.4052 - auc: 0.7351 - val_binary_crossentropy: 0.4080 - val_auc: 0.722
Epoch 19/30
13s - loss: 0.4050 - binary_crossentropy: 0.4050 - auc: 0.7355 - val_binary_crossentropy: 0.4079 - val_auc: 0.723
Epoch 20/30
14s - loss: 0.4047 - binary_crossentropy: 0.4047 - auc: 0.7359 - val_binary_crossentropy: 0.4078 - val_auc: 0.723
Epoch 21/30
14s - loss: 0.4045 - binary_crossentropy: 0.4045 - auc: 0.7366 - val_binary_crossentropy: 0.4077 - val_auc: 0.723
Epoch 22/30
14s - loss: 0.4044 - binary_crossentropy: 0.4044 - auc: 0.7369 - val_binary_crossentropy: 0.4076 - val_auc: 0.723
Epoch 23/30
14s - loss: 0.4042 - binary_crossentropy: 0.4042 - auc: 0.7371 - val_binary_crossentropy: 0.4075 - val_auc: 0.723
Epoch 24/30
14s - loss: 0.4040 - binary_crossentropy: 0.4040 - auc: 0.7375 - val_binary_crossentropy: 0.4073 - val_auc: 0.724
Epoch 25/30
14s - loss: 0.4038 - binary_crossentropy: 0.4038 - auc: 0.7377 - val_binary_crossentropy: 0.4072 - val_auc: 0.724
Epoch 26/30
13s - loss: 0.4037 - binary_crossentropy: 0.4037 - auc: 0.7380 - val_binary_crossentropy: 0.4071 - val_auc: 0.724
Epoch 27/30
14s - loss: 0.4035 - binary_crossentropy: 0.4035 - auc: 0.7382 - val_binary_crossentropy: 0.4070 - val_auc: 0.724
Epoch 28/30

```

```

train_loss = history.history['loss']
epochs = range(1, len(train_loss) + 1)

```

```

plt.plot(epochs, train_loss, 'b', label='Training loss')
plt.title('Training loss')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.legend()
plt.grid(True)
plt.show()

```

