

MLP Report (Week 7)

This week we examined on multiple months prediction, and apply entropy evaluation following the algorithm:

```
# Evaluation by log probability with base `e`
# Range: (0, ln(num_classes))
def entropy(y_pred_proba, y_test):
    true_probs = np.sum(y_pred_proba * y_test, axis=1)
    log_probs = np.empty_like(true_probs)
    for i, p in enumerate(true_probs):
        if p == 0:
            print(f"Sample {i}: True class probability is 0, setting log to NaN")
            log_probs[i] = np.nan
        else:
            log_probs[i] = np.log(p)

    entropy = -np.nanmean(log_probs)
    return entropy
```

Within classes from **0 to 6+**, entropy ranges from **0 to 2.079**. The results below are all based on predicting delinquency status after 3 months.

Iteration Method

Under this methods, two different approaches are applied. In the following algorithm, either pass original `pred_proba` to `input`, or encode into one-hot form to align with training data.

```
# iterate to predict status after n months
def predict_n_months(mlp, n, input):
    for _ in range(n):
        pred_proba = mlp.predict_proba(input)

        # choose one following method
        # 1. remain probability form
        input = pred_proba

        # 2. encode to one-hot
        input_labels = np.argmax(pred_proba, axis=1)
        input = np.zeros_like(pred_proba)
        input[np.arange(len(input)), input_labels] = 1

    return pred_proba

y_pred_proba = predict_n_months(mlp, MONTH_AHEAD, X_test)
```

Generally, the latter achieves better result.

1. remain probability form

average probability = 0.12584506748065927
entropy = 0.16520659321309436
brier score = 0.037273658636576584
adjusted brier score = 0.0027203744763043556

With Credit Score:

average probability = 0.13524271686443104
entropy = 0.12263973142272337
brier score = 0.03404474790524376
adjusted brier score = 0.002073965841030571

2. encode pred_proba to one-hot

average probability = 0.24346980135274554
entropy = 0.09765431522867496
brier score = 0.031637665749323626
adjusted brier score = 0.00206502130955303

With Credit Score:

average probability = 0.2452789950399672
entropy = 0.10265724155852532
brier score = 0.031817913828540344 adjusted brier score = 0.002108563723060312

With 5 main features:

average probability = 0.2501087228409682
entropy = 0.10682508970748208
brier score = 0.03168545357708819
adjusted brier score = 0.0020835252409971633

Train MLP with status 3 months ago

We can also modify the training process according the months to predict ahead, as:

```
MONTH_AHEAD = 3
# in function preprocess()
df['Next Loan Delinquency Status'] = df.groupby('Loan Sequence Number')['Current
Loan Delinquency Status'].shift(-MONTH_AHEAD)
```

average probability = 0.30554111135043394
entropy = 0.0793874576832437
brier score = 0.02810046936215188
adjusted brier score = 0.001936038324794717

With Credit Score:

average probability = 0.31309216414193974
entropy = 0.07540912794051383
brier score = 0.02790038888099853
adjusted brier score = 0.0019538893776107928

With 5 main features:

average probability = 0.3154518015108996

entropy = 0.07496824565015968

brier score = 0.027976407199123924

adjusted brier score = 0.0019539852498494854