Q2

b) Using 5-fold cross validation, the result is as follows:

```
5-fold validation performance on binary dataset (epoch: mistakes):
{5: 25.4, 10: 22.0, 20: 19.0, 30: 22.0, 50: 20.4, 100: 20.6, 200: 20.0}
5-fold validation performance on count dataset (epoch: mistakes):
{5: 51.8, 10: 52.0, 20: 36.4, 30: 53.0, 50: 33.0, 100: 26.6, 200: 24.6}
```

Therefore, the optimal number of epochs for binary dataset is 20, and for count dataset is 200.

Results of training the model on optimal number of epochs respectively:

```
Results of training on binary dataset with 20 epochs:
{'Train': 3, 'Test': 27}
Results of training on count dataset with 200 epochs:
{'Train': 4, 'Test': 30}
```

c) For binary dataset:

```
words with max 15 weights:
['client' 'renam' 'guess' 'sign' 'wed' 'hous' 'thi' 'pleasur' 'find'
 'simpli' 'out' 'amaz' 'pain' 'oppos' 'debat']
words with min 15 weights:
['www' 'us' 'vipul' 'the' 'sport' 'unusu' 'onc' 'whether' 'numberx' 'star'
 'gari' 'bulk' 'reach' 'cash' 'clearli']
```

For count dataset:

```
words with max 15 weights:
['renam' 'statement' 'dead' 'numbercnumb' 'william' 'out' 'guess' 'domain'
 'numberam' 'board' 'environ' 'profil' 'yesterdai' 'pleasur' 'int']
words with min 15 weights:
['numberpm' 'filenam' 'button' 'reach' 'settl' 'cnn' 'us' 'www' 'dave'
 'never' 'sale' 'offici' 'writer' 'http' 'technic']
```

Q3

a) For Naïve Bayes using MultinomialNB:

```
Naive Bayes trained on binary dataset:
{'train': 115, 'test': 34}
Naive Bayes trained on count dataset:
{'train': 133, 'test': 39}
```

b) For Logistic Regression:

```
Logistic Regression trained on binary dataset:
{'train': 3, 'test': 24}
Logistic Regression trained on count dataset:
{'train': 3, 'test': 32}
```