Q1. *Emotion detection in music* as a well posed learning problem:

T: classifying music with respect to the type of emotion like happy, sad, relaxing etc.

P: percent of songs (parts of songs) correctly classified

E: songs (parts of songs) together with human annotated emotion labels

Q3.

In the case of underfitting, a simple model is fit to data (*training set error is high*), *bias (validation set error) is high*. When the training set size gets larger, as model still do not fit to data well (*training set error gets higher*), model points have low variance and *relatively lower bias (validation set error)*.

When it comes to overfitting, a complex model that explains data well has high variance and *low training set error*. However, *the validation set error is high* as well, as data points include some noise, too. Whenever more training data are provided, the variance of model data points gets lower and *training set error is increased*. On the other hand, *validation set error decreases*, which means the gap between training set error and validation set error gets smaller in other words errors start to converge.

4.

$$P(S \mid E, H) = \frac{P(E, H \mid S)\ P(S)}{P(E, H \mid S)\ P(S) + P(E, H \mid NS)\ P(NS)}$$

independence assumption ↓

$$= \frac{P(E \mid S)\ P(H \mid S)\ P(S)}{P(E \mid S)\ P(H \mid S)\ P(S) + P(E \mid NS)\ P(H \mid NS)\ P(NS)}$$

$$= \frac{1500/10000 \quad 800/10000 \quad 10000/15000}{\frac{1500}{10000}\ \frac{800}{10000}\ \frac{10000}{15000} + \frac{20}{5000}\ \frac{200}{5000}\ \frac{5000}{15000}}$$

$$\cong 0.99$$

As $P(S \mid E, H) \cong 0.99 > 0.9$,

the message will be rejected as spam.

Q5.

Imagine a multi-class classification problem where the true membership is as follows:

Real: [1 0 0 0 0]

Let the system generates a prediction as a result of the forward propagation as below:

Prediction: [0.1 0.5 0.1 0.1 0.2]

In the case of cross entropy, loss simplifies to

$$H(\hat{y}, y) = -y_i \log(\hat{y}_i) = -(1.\log 0.1) \cong 2.303$$

As for misclassification loss, the system counts it just as a misclassification, which does not take into consideration the individual values in the prediction vector. It cannot make a difference between the prediction vectors that are given below:

[0.1 0.5 0.1 0.1 0.2], [0.34 0.36 0.1 0.1 0.1]

When it comes to MSE, the calculation pays attention to all values in the prediction vectors. The overall result can be attributed to the diversions from the incorrect membership cell values. Thus, misclassification loss barely calculates the error, MSE is quite sensitive to incorrect values, cross entropy is in between.