Answer the questions in 90 minutes.

1. **(20 points)** Well posed learning problems can be defined as follows:

   A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

   Please state T, P, and E in *emotion detection in music* learning problem.

2. **(20 points)** Covariance between two variables can be defined by the following formula:

$$Cov(X, Y) = E((X - E(X))(Y - E(Y)))$$

   where E is the expected value.

   Show that $Cov(X, Y) = 0$ if X and Y are independent random variables.

   Describe the relationship between two variables if their covariance is positive/negative.

3. **(20 points)** Please explain underfitting and overfitting with respect to training set size and model complexity.

4. **(20 points)** Suppose that a Bayesian spam filter is trained on a set of 10000 spam messages and 5000 messages that are not spam. The word "enhancement" appears in 1500 spam messages and 20 messages that are not spam, while the word "herbal" appears in 800 spam messages and 200 messages that are not spam. Estimate the probability that a received message containing both the words "enhancement" and "herbal" is spam. Will the message be rejected as spam if the threshold for rejecting spam is 0.9?

5. **(20 points)** When a neural network is used to perform classification, cross-entropy loss is preferred over misclassification loss and mean squared error. Please provide appropriate reasoning behind this.

$$\text{Misclassification Loss:} \qquad l(\theta) = \frac{1}{N}\sum_{i=1}^{N} \delta(h_\theta(x^{(i)}) \neq y^{(i)}) \tag{1}$$

$$\text{Mean Squared Error:} \qquad l(\theta) = \frac{1}{N}\sum_{i=1}^{N} (h_\theta(x^{(i)}) - y^{(i)})^2 \tag{2}$$

$$\text{Cross-entropy Loss:} \qquad l(\theta) = \sum_{i=1}^{N} y^{(i)} log(h_\theta(x^{(i)})) + (1 - y^{(i)}) log(1 - h_\theta(x^{(i)})) \tag{3}$$