Assignment1
Kewei WANG

1. Question1:
   (a) It's not machine learning. Because it just tells the machine how to distinguish between prime or non-primed number, but learning systems are not directly programmed to solve a problem, instead develop their own program based on examples and experiences.
   (b) It's machine learning. Because this is the learning system and it's developed their own program based on examples and experiences, and it can develop themselves to predict things.
   (c) It's not machine learning. Because it is just a physical problem, and the exact result can be obtained by some physical formulas.
   (d) It's not machine learning. It is the optimization.

   Question2:
   Supervised classification learning. The goal is to decide whether an article is about science policy. Input data: the newspaper articles, especially the words contained in these articles. Feature and model selection: the times of the words ——"science", "public", "open", "university", "government", "funding", "education", "justice", "law" appearing in the article. If these words' appearance time reach a specific level, the article is about science policy, otherwise it's not. Training: Use several science policy articles and several other article to decide the specific level of the times of the words mentioned before. Test: Use new articles to test the model. If it's exact enough, the model will be applied to predict whether the article is about science policy.

   Question3:
   (a) It's not machine learning problem.
   (b) It's a machine learning problem. The type is supervised prediction. The design matrix is the house's building plan and a civil engineer's records of plans and heating loads for houses in the same neighborhood. The target vector is the house's heating load (i.e., the amount of energy that is needed to maintain the temperature).

2. Question4:
   (a) Because $(MM^T)^T = M^TM$, it's symmetric.
   Because M is m*n matrix, M^T is n*m matrix, MM^T is m*m matrix and it's square and so is M^TM.
   If all the number in the M matrix is real number, the two matrixes in the question will be real.
   (b) According to $(MM^T – a*E) * v1 = 0$, $(M^TM – a*E) * v2 = 0$, where a is eigenvalue, v1, v2 are eigenvectors, the $dia(MM^T – a*E) = dia(M^TM – a*E)$, since a is equal.

Because $(MM^T - a*E)^T = (M^TM - a*E)^T$ (it's symmetric), v1 = v2. They will also have the equal eigenvectors.

(c) $MM^T = U\Sigma^2U^T$, $M^TM = V\Sigma^2V^T$

(d) Because $M^TM = V\Sigma^2V^T$, and $M = U\Sigma V^T$, MM^T 's Σ is square of M's, the singular value will be the square root of eigenvalues.

Question5:

(a) $\begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}$

The covariance matrix is C = 1/3 * $\begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}$ * $\begin{pmatrix} -1 & -1 \\ 0 & 0 \\ 1 & 1 \end{pmatrix}$ = $\begin{pmatrix} 2/3 & 2/3 \\ 2/3 & 2/3 \end{pmatrix}$

The eigenvalues are a1 = 4/3, a2 = 0,

Then the eigenvectors are $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$,

So, P = $\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$, the first principal axis is $(1/\sqrt{2} \quad 1/\sqrt{2})$

(b) Y = $(1/\sqrt{2} \quad 1/\sqrt{2})$ * $\begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}$ = $(-\sqrt{2}\ 0\ \sqrt{2})$

(c) Var = 4/3

3. Question6:

(a) Acc1 = (100 + 97)/(100 + 97 + 3) = 0.985
Acc2 = (100 + 96)/200 = 0.98

(b) Algorithm2 is better, although Algorithm1 has better accuracy, when it comes to the task of identifying poisonous mushrooms, it's better to precision or False discovery rate between two algorithms.
PPV1 = 0.97
PPV2 = 1
FDR1 = 0.03
FDR2 = 0
So the Algorithm2 has better PPV while less FDR, it will work better in the task of identifying poisonous mushrooms.

Question7:

	Labeled normal	Labeled fraud
normal	TN	FN
fraud	FP	TP

As we can see from the sheet above:
[a]must be minimized. False positive rate means the transaction is fraud but hasn't been detected as fraud. Because undetected frauds are quite costly to the bank, compared to establishing that a transaction was, in fact, not fraudulent, minimizing the rate of false positive will be the most beneficial to the bank.

4. Question8:

(a) $J_{RR}(\theta) = (y - X * \theta)^T * (y - X * \theta) + \lambda * \theta^2$

Let the derivative of $J_{RR}(\theta)$ equal to 0.

$$2 * X^T * X * \theta - 2 * X^T * y + 2 * \lambda * \theta = 0$$

$$\theta = (X^T * X + \lambda * I)^{-1} * X^T * y$$

(b) Because when training the training dataset, if there is overfitting, least-squares regression can't prevent it. However, the ridge regression estimator add a regularization term which will help penalize model complexity during training. Thus the final training model will be prevented from overfitting.