

Exercise for Introduction to Machine Learning,
Advanced materials processing with intelligent systems

As seen in class, one of the most practical definition of machine learning is the following:

*Def := 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**.*

Suppose your email program checks which emails you do mark as spam, and based on that, learns how to filter spam better. What is task T in this setting?

1. Classifying emails as spam or not spam
2. Watching you label emails as spam or not spam
3. The number of emails correctly classified as spam
4. None of the above, this is not a machine learning problem

Classify the following entries in these categories (E, T, P):

Topic: playing checkers.

1. The experience of playing many games of checkers
2. Playing checkers.
3. The probability that the program will win the next game.

You're running a company, and you want to develop algorithms to address each of these problems:

- 1) You have a large inventory of identical items. You want to predict how many of these items will be sold over the next three months
- 2) You would like to have a software that examines individual customers' accounts, and for each account, decide if it has been compromised.

Should these two issues be treated as a classification or regression problems?

1. Treat both as classification problems
2. Treat problem 1 as a classification problem, problem 2 as a regression problem
3. Treat problem 1 as a regression problem, problem 2 as a classification problem
4. Treat both as regression problems

Which of the following problems should be addressed using an unsupervised learning algorithm?

1. Given email labeled as spam/non-spam, learn a spam filter
2. Given a set of news articles, group them into sets of articles about the same story
3. Given a dataset of patients diagnosed as either having diabetes or not, learn to classify new patients as having diabetes or not.

Given a collection of raw signals from the optical emission from the process zone in a window of 20 ms, and the corresponding labels in terms of defect content.

Which of the following algorithms should I use:

1. Artificial Neural Network
2. Logistic regression
3. Linear Regression