# Machine Learning, Fall 2018: Project 2

Your Name Here

You may use any programming language you like (Matlab, C++, C, Java... ). All programming must be done from first principles. You are only permitted to use existing tools for simple linear algebra such as matrix multiplication/inversion. **Do NOT use any toolkit that performs machine learning functions.**

## 1 Extra credit: Parameter Estimation

If $X$ (e.g. packet arrival density) is Poisson distributed, then it has pmf

$$P(X|\lambda) = \frac{\lambda^X e^{-\lambda}}{X!}$$

### 1.1 MLE and MAP estimates

1. Show that $\hat{\lambda} = \frac{1}{n}\sum_i X_i$ is the maximum likelihood estimate of $\lambda$ and that it is unbiased (that is, show that $\mathbb{E}[\hat{\lambda}] - \lambda = 0$).
2. Recall that the Gamma distribution has pdf:

   $$p(\lambda|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}, \quad \lambda > 0$$

   Assuming that $\lambda$ is distributed according to $\Gamma(\lambda|\alpha,\beta)$, compute the posterior distribution over $\lambda$.
3. Derive an analytic expression for the maximum a posteriori (MAP) estimate of $\lambda$ under a $\Gamma(\alpha,\beta)$ prior.

## 2 Decision Trees

### 2.1 ID3

Consider the following set of training examples for the unknown target function $< X_1, X_2 > \rightarrow Y$.

| $Y$ | $X_1$ | $X_2$ | Count |
|---|---|---|---|
| + | T | T | 3 |
| + | T | F | 4 |
| + | F | T | 4 |
| + | F | F | 1 |
| - | T | T | 0 |
| - | T | F | 1 |
| - | F | T | 3 |
| - | F | F | 5 |

1. What is the sample entropy $H(Y)$ for this training data (with logarithms base 2)?
2. What are the information gains $IG(X_1) \equiv H(Y) - H(Y|X_1)$ and $IG(X_2) \equiv H(Y) - H(Y|X_2)$ for this sample of training data?
3. Draw the decision tree that would be learned by ID3 (without postpruning) from this sample of training data.

# 3 Perceptron

## 3.1 Liver disease classification

This data set analyzes some liver disorders that might arise from excessive alcohol consumption (the first 5 features), and the number of half-pint equivalents of alcoholic beverages drunk per day for each individual. The task is to select if a given individual suffers from alcoholism.

Evaluate your Perceptron implementation on the Bupa Liver Disorder dataset that is linked on the syllabus. The classification problem is to predict whether an individual has a liver disorder (indicated by the selector feature) based on the results of a number of blood tests and levels of alcohol consumption. Use 90% of the dataset for training and 10% for testing. Average your results over 50 random splits of the data into training sets and test sets.

1. What happens when the learning rate is large vs small?
2. Does the algorithm converge? Plot the instances in a way that explains your answer.
3. Come up with a confidence metric in your classification. (For example come up with an activation function that might correspond to confidence.) Create a scatter plot for confidence vs classification result for all test instances.
4. Devise a new synthetic feature that would increase accuracy. Explain why the accuracy should increase in a quantitative manner. Run your algorithm again and report improved accuracy.