1 Dimensional Reduction

- Compute the first three principal components of the data in dataset `https://archive.ics.uci.edu/dataset/109/wine`, report the variance.
- Derive the equation $S_B w = \lambda S_W w$ from minimizing the following function $J(w) = \frac{w^T S_B w}{w^T S_W w}$, which is the main objective in the LDA dimensional reduction method.

2 Model Evaluation

- Compute the optimal decision for a model $p(y|x)$, if the loss function is chosen as $L(y, a) = |y - a|$ where a is the action, y is the true value.
- Use the cross validation method to evaluate your regression model. The data is in the link `https://archive.ics.uci.edu/dataset/9/auto+mpg`, use the ridge regression. Determine the best hyperparameters from the cross-validation.
- Using the random forest and gradient boosting method (you might need to download the XGBoost package separately) to do the classification for the dataset `https://archive.ics.uci.edu/dataset/222/bank+marketing`, report your parameters and compare your results.
- Compute the evidence for the following two models and use the result to compare the models. H_0 is a uniform distribution, and the probability is

$$p(x|H_0) = \frac{1}{2}, \quad x \in (-1, 1)$$

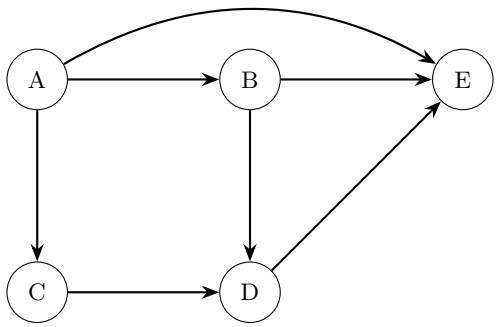
The model H_1 is a nonuniform distribution with an unknown parameter $m \in (-1, 1)$:

$$p(x|m, H_1) = \frac{1}{2}(1 + mx), \quad x \in (-1, 1)$$

Given the data $D = (0.3, 0.5, 0.7, 0.8, 0.9)$, compute the evidence H_0 and H_1.

3 Graph Probability Model and K means

- The random variables are $x_1, \ldots, x_n$. The conditional probability of x_i only depends on two previous variables x_{i-1}, x_{i-2}. Write down the graph for this model.
- Using d-separation to verify the conditional independence of:
  (a) $p(a, d|c)$
  (b) $p(a, e|b)$
  (c) $p(a, d|b)$
  (d) $p((a, c), d|b)$



- Using the K-means method to do the clustering for the data in `https://archive.ics.uci.edu/dataset/352/online+retail` (Choose proper features, do not use all of them). Find the best K using the two methods discussed in the lecture.