Supervised

1. Decision Tree
   `PSEUDO CODE HERE with O(n) on steps`
2. Naive Bayes
   `PSEUDO CODE HERE with O(n) on steps`
3. KNN
   - Non-Parametric
   - fit cost: store $O(nd)$
   - Prediction Cost: $O(nd)$?
   - Usage:

   `PSEUDO CODE HERE with O(n) on steps`
4. Linear Regression
5. Non-linear Regression (Supervised)

Unsupervised

1. K-Means (Unsupervised) `PSEUDO CODE HERE with O(n) on steps`
2. DBSCAN (Unsupervised)
3. Outlier Detection Methods
4. Ensemble Methods
5. Cross-Validation `PSEUDO CODE HERE with O(n) on steps`
6. Definitions
7. Linear Algebra Notes Basics
   (a) $w^T x_i = \sum_{j=1}^{d} w_j x_{ij}$, x_i, w is d x 1
   (b) $a^T A b = b^T B^T a$ both sides are vectors
   (c) $\frac{1}{2}\|Xw - y\|_2^2 = \frac{1}{2}\sum_{i=1}^{n}(w^T x_i - y_i) = \frac{1}{2} w^T X^T X w - w^T X^T y + \frac{1}{2} y^T y$
   (d) $\nabla \text{const} = 0, \nabla w^t b = w, \nabla \frac{1}{2} w^T A w = Aw$ if A symmetric
   (e) $\nabla \frac{1}{2}\|Xw - y\|_2^2 = X^T X w - X^T y$
   (f) Normal equation $X^T X w = X^T y$
   (g) $(Xw - y)^T V (Xw - y) = \sum_{i=1}^{n} v_i (w^T x_i - y_i)^2$

   Run Time
   (a) $X^T y : O(nd)$
   (b) $X^T X : O(nd^2)$
   (c) solve d x d system of equations : $O(d^3)$
   (d) solve normal equation : $O(d^3 + nd^2)$

   Gradient Descent
   (a) $w^{t+1} = w^t - \alpha^t \nabla f(w^t) = w^t - X^T(Xw^t - y)$ (least square)
   (b) cost $O(nd)$ no need to form $X^T X$
   (c) total cost $O(ndt)$
   (d) faster for large d, works generally
8. Multivariable Calc Notes
9. Probability Notes