Algorithms

MinSeok Song

Spectral clustering

Definition 1. *$Ncut(A, B)$ is defined as*

$$Ncut(A, B) = cut(A, B)(\frac{1}{d(A)} + \frac{1}{d(B)})$$

where $d(A) = \sum_{i \in A} d_i$.

Remark 1.
- The intuition is that when A is relatively small, the $\frac{1}{d(A)}$ will be large, hence discouraging the isolating small groups.
- Finding minimum Ncut is equivalent to finding vector v that minimizes

  $$\frac{v^T L v}{v^t D v} \text{such that } v^t D1 = 0, v_i \in \{a, b\}$$

  where $L = D - W$.

When the data is separable linearly

There are several ways. First, linear programming.

Algorithm 1 Linear Programming for Classifier

$$\begin{aligned} \textbf{Objective:} \quad & \text{minimize} \quad \mathbf{u} = (0, \dots, 0) \quad \text{(dummy variable)} \\ \textbf{Subject to:} \quad & A\mathbf{w} \geqslant \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \\ & A_{i,j} = y_i x_{i,j} \quad \text{(where } j\text{'th element of the vector } x_i) \end{aligned}$$

Remark 2.
- u is a dummy variable here; we essentially only check if the constraint is satisfied.
- This is only applied for when the data is separable.
- We can formula the regression problem with loss function $l(h, (x, y)) = |h(x) = y|$ using linear programming.

Algorithm 2 Batch Perceptron

function BatchPerceptron(x_1, y_1, ..., x_m, y_m)
  $w(1) \leftarrow (0, \ldots, 0)$
  for $t \leftarrow 1, 2, \ldots$ **do**
    if there exists i such that $y_i\langle w(t), x_i\rangle \leqslant 0$ **then**
      $w(t+1) \leftarrow w(t) + y_i x_i$
    else
      output $w(t)$
      break
    end if
  end for
end function

Remark 3.

- Note that $y_i(w^{(t+1)}, x_i) = y_i(w^{(t)}, x_i) + \|x_i\|^2$
- The algorithm must stop after at most $(RB)^2$ iterations, where $R = \max_i \|x_i\|$ represents a data spread, and $B = \min\{\|w\| : i \in [m], y_i\langle w, x_i\rangle \geqslant 1\}$ represents margin.
- To prove this, it suffices to show that $1 \geqslant \frac{\langle w*, w^{(T+1)}\rangle}{\|w*\|\|w^{(T+1)}\|} \geqslant \frac{\sqrt{T}}{RB}$, which we proceed by bounding numerator and denominator separately.
- We can prove that this bound is tight. For some vector $w^* \in \mathbb{R}^d$, the algorithm incurs $m = (BR)^2$ errors (considering $m = d$).
- Moreover, for $d = 3$, an algorithm can be designed to commit exactly (m) errors for any given $m \in \mathbb{N}$, serving as an upper bound concurrently

Logistic Regression

Definition 2. *Fit the logistic function* $\phi_{sig}(x) = \frac{1}{1+exp(-\langle w,x\rangle)}$ *with minimization scheme* $w = argmin_{w\in\mathbb{R}^d} \frac{1}{m}\sum_{i=1}^{m} \log(1 + exp(-y_i\langle w, x_i\rangle))$.

Remark 4.

- The explaratory variable is between 0 and 1, making it interpretable as a probability.
- Appropriate for binary classification.
- Logistic loss function is a convex function so it's efficient to minimize.