Quiz 6 Cheat Sheet

Kyle Kazemirian
CS 3190
11/29/20

Clustering: Input $X = \{x_1, x_2, \ldots, x_n\}$ $X \subset \mathbb{R}^d$
Distance $D: X \times X \to \mathbb{R}^+$ $D(x_1, x_2) = \|x_1 - x_2\|$
Goal: k subsets $\{x_1, x_2, \ldots, x_k\}$ $x_i \in X$

$$\phi_S(x) = \underset{s_i \in S}{\text{argmin}} \|x - s_i\|$$

$$\sum_i r_i^2 = \sum_{i=1}^{n} \|x_i - \phi_S(x_i)\|^2$$

Lloyd's Algorithm: $\text{Cost}(X, S) = \sum_{x \in X} \|\phi_S(x) - x\|^2$

(1) Choose k points $S \subset X$ (2) $\forall x \in X$, assign x to X_i so $\phi_S(x) = s_i$ (3) $\forall s_i \in S$ update $s_i = \frac{1}{|X_i|} \sum_{x \in X_i} x$ (4) until S is unchanged.

Mixture of Gaussians covariance matrix:

$$\Sigma_i = \sum_{x \in X_i} (x - \mu_i)(x - \mu_i)^T$$

Loss function $f(\alpha) = \mathcal{L}(g_\alpha, (X, y)) =$
$= \sum_{i=1}^{n} f_i(\alpha)$ where $f_i(\alpha) = \ell(z_i = y_i g_\alpha(x_i))$

Perceptron Algorithm: Initialize $w = y_i x_i$ for any i in (X, y)
repeat:
$\forall (x_i, y_i)$ s.t. $y_i \langle x_i, w \rangle < 0$; (misclassified)
update $w \leftarrow w + y_i x_i$
until T steps or no more misclassifications.
return $w \leftarrow \frac{w}{\|w\|}$