Crib 4

04 Weighted, Total Least Squares

by Alvin Wan . alvinwan.com/cs189/fa17

Note that in the objective functions below, you may choose to featurize your data i.e., replace all x_i with $\phi(x_i)$

1 Weighted Least Squares

1. Objective: $\min_w \sum_{i=1}^n \omega_i(x_i^T w - y_i)$ for $\omega_i, y_i \in \mathbb{R}, x_i, w \in \mathbb{R}^d$
2. Motivation: Each sample has a different level of "importance", so we weight them each differently.
3. Probabilistic interpretation: $y = \phi(x_i)^T w + z_i$, where the noise in our labels is not uniform across samples.

2 Total Least Squares

1. Objective: $\min \| \begin{bmatrix} \tilde{X} & \hat{y} \end{bmatrix} \|_F^2$ such that $y + \tilde{y} = (X + \tilde{X})w$
2. Motivation: Make more robust, by minimally perturbing the data ($\tilde{X}$) to lie within the minimally perturbed ($\tilde{y}$) space.
3. Probabilistic interpretation: Both X and y experience some degree of noise.
4. Total Least Squares (TLS) is equivalent to Principal Component Analysis (PCA)
5. Eckhard-Young Theorem: Consider A and its SVD: $A = U\Sigma V^T$. Take a k-rank approximation $A_k = \sum_{i=1}^k \sigma_i u_i v_i^T$, then $\|A - A_k\| \leq \|A - \tilde{A}\|_F$ for all $\tilde{A}$ with rank $\leq k$.