RECITATION

10/21/2022

Questions …

Q1. implement the SGD.

Q2. compare SGD with BGD.

Q3. implement the Ridge Regression with Normal Equation.

Q4. implement the LWLR.

Q5. explore the function of τ.

Q6. apply linear regression on Pandemic data analysis.

Mistakes

- SGD: epoch number? Expression?
- LWLR: Expression?
- More?

$$\theta_j \mathrel{+}= \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}$$

$$\theta_{RR}^{*} = (X^T X + \alpha I)^{-1} X^T y$$

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{2\times 2}$$

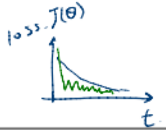
BGD VS SGD

Batch gradient descent

Repeat until convergence{
$\theta_j = \theta_j + \alpha \sum_{i=1}^{m} (y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)}$ for every j
}

Stochastic gradient descent

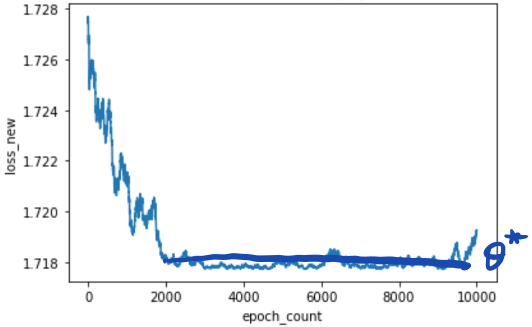
Repeat until convergence{
for $i = 1 \dots m$ {
$\theta_j = \theta_j + \alpha(y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)}$ for every j
}
}



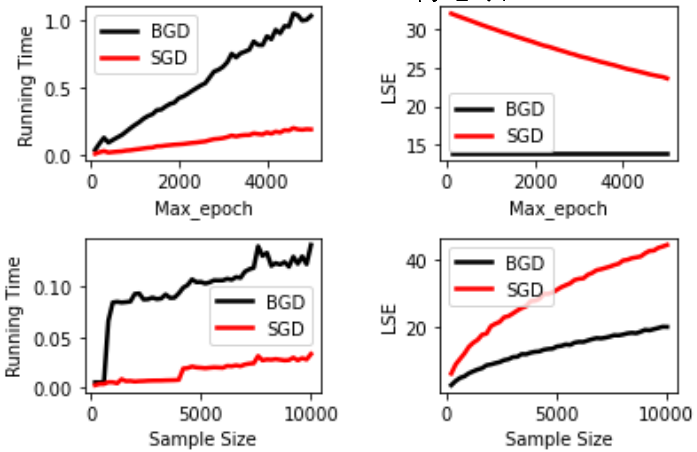
θ is updated each time a training example is read

- Stochastic gradient descent gets θ close to minimum much faster
- Good for regression on large data

2021215267 谢龙



2022213976 陈思琪



Locally Weighted Linear Regression

Motivation: Assumptions are true in 'local'.

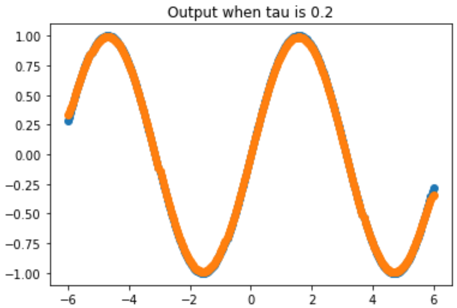
I. Curve Fitting or Prediction?
II. Process the bounded input data?
   I. Ignore
   II. Append dummy values
   III. Padding with 0s
   IV. etc
III. Analytical or Numerical?

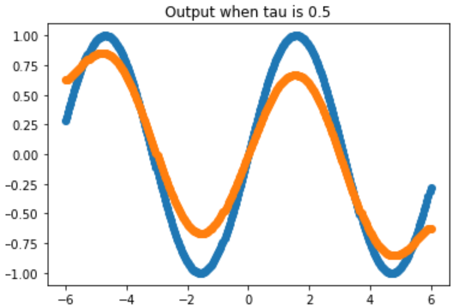
$$W = diag\left(\begin{vmatrix} \omega_0 \\ \cdots\cdots \\ \omega_{win\_len} \end{vmatrix}\right) \qquad LS(\Theta) = \sum_i^n \omega_i (y_i - \Theta_i^T{}_i x_i)^2 = (X\Theta - Y)^T W (X\Theta - Y) = \theta^T X^T W X \theta - 2Y^T W X \theta + Y^T W Y$$

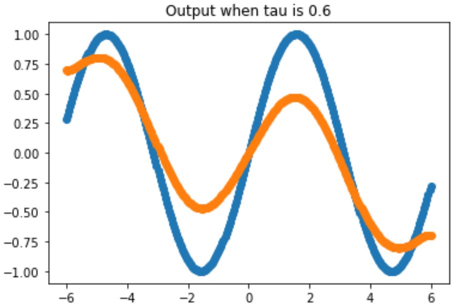
Exploring τ

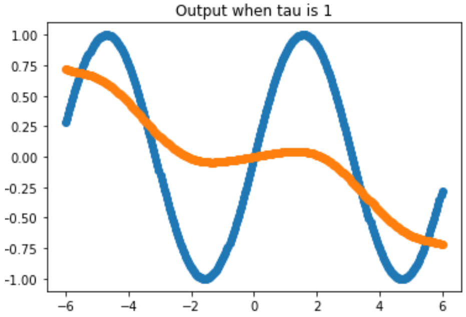
Parameter τ controls the variance of the weight over all selected input data.

2021215267 谢龙









- **When $\tau \to \infty$, all $\omega_i \to 1$. It degenerates to normal linear regression;**
- **When $\tau \to 0$, all $\omega_i = \P\{i = i_{target}\}$. The model fits the target point itself.**

$$\omega^{(i)} = exp\left(- \frac{(x^{(i)} - x)^2}{2\tau^2} \right)$$

Pandemic Data Analysis

Various methods to make predictions:

- Directly using one slide window. 45 - 50 → 51 52 53 54 55
- Rolling the slide window forward after predicting one time unit.

45-50 → 51

46-50+51 → 52