Lecture 3 | Locally Weighted & Logistic Regression

Date de création	@December 19, 2024 4:46 PM
Étiquettes	

Outline

- Locally Weighted Regression
- Probabilistic Interpretation
- Logistic Regression
- Newton's Method

Recap

(x^i, y^i) → ith exemple

$x^{(:)} \varepsilon \mathbb{R}^{n+1}, y^i \varepsilon \mathbb{R}, xo = 1$

m = exemples, n = features

$ho(x) = \Sigma(j,n)\Theta jxj = \Theta^T x$

$J\Theta = \frac{1}{2}\Sigma(i,m)((ho(x^i) - y^i))^2$

Intro

Parametric Learning

> implement weighted regression

Let's say. For Linear Regression is fit Θ to your data and return $\Theta^t x$. You should minimazie Loss with this Θ.

Locally weighted regression (non-parametric algorithm)

Fit Θ to minimize a modifield loss function: $\sum w^i (y^i \Theta^{t^i} x)^2$, where$w^i = exp(-\frac{(x^i - x)^2}{2\tau^2})$

propety 1: if $|x^i - x|$ is small, w^i ~ 1

propety 2: if $|x^i - x|$ is large, w^i ~ 0

Supose we compute the house price, like that: y^i = $\Theta^t x^i + \epsilon^i$, where this ϵ^i is some error.

This simple mean that. the value of the house is our hypothesis but some "force" of nature or noyse, as you wish.

IID → The error of one house doesn't affect another.

Logit

One reason we use Linear regression is that, we want our $h(x)$ ϵ [0, 1].

$h\theta(x) = g(\theta^y x) = \frac{1}{1+e^{-\theta x}}$, we call g(z) = $\frac{1}{1+e^{-\theta x}}$, 'sigmoid function' or logistic function.

Binary Problem

$P(y|x;\theta) = h(x)^y (1 - h(x))^{1-y})$

"Log Likelihood"
$l(\theta) = logL(\Theta)$

= sum(i=i, n) $y^i logh\theta(x^i) + (1 - y^i)log(1 - h\theta(x^i)$

Supose you have a data (x, y). You compute the Log of $L(\theta)$, then you choose θ to maximize $L(\theta)$ using gradient ascent. Ascent means we want the maximum value:

$\Theta j := \Theta j + \alpha \frac{\partial}{\partial \Theta j} * l(\Theta)$

Newton's Method

In comparison with Gradient Ascent or Descent you can give high steps using Newton's method.

Ex: Have F, want to find Θ such that $F(\Theta) = 0$

Newton's method uses the concept of quadratic interations, which make the convergence much more faster.

When Θ is a **vector**:
$\Theta^{(t+i)} := \Theta^{(t)} + H^{-1}\nabla \theta l$, where H is The Hessian Matrix.

Newton's method is faster than Gradient Ascent or Descent, but if you have high dimensional data the vector increases the coast of each interration.

Questions

- When use Linear weighnted regression?

  a few features, lots of data and don't care about what features to select.

- Why specificly choose the Sigmoid Function?
- Why we use squared error?
- if that data we want to fit isn't in the format $\Theta o + \Theta x?$ but $\Theta o + \Theta xi + \Theta iix^2?$

- What's Locally weihted regression? And why does it matter?

  It consist in select some area to make predictions based on that area (K-Neighbor)

- What's the difference between parametric algorithms and non-parametric algorithms.

  Parametric: Fit fixed set of parameters Θi, to data.

  Non-Parametric: The number of parameters grows with data size.

- What's the Hessian Matrix? What does it matter?

  A vector of **partial derivatives**:

  $\frac{\partial^2 l}{\partial \Theta i, \partial \Theta j}$

Futher Reding

ML | Locally weighted Linear Regression - GeeksforGeeks

A Computer Science portal for geeks. It contains well written, well thought and well explained computer science and programming articles, quizzes and practice/competitive

https://www.geeksforgeeks.org/ml-locally-weighted-linear-regression/