ML Assignment 1

Author : Aman Bhatia

Dated : 07 Feb 2016

Linear Regression

(a)

- Learning Rate : 1.0
- Stopping Criteria : If $\forall j$, $\theta_j < \varepsilon$ where $\varepsilon = 10^{-8}$
- # Iteration to converge : 266
- Final set of parameters : $\theta_0 = 5.83913505$
  $\theta_1 = 17.68267541$

(e)

- $\eta = 0.1$: Learning is very slow, it takes 2406 iterations to converge, steps in the contour diagram are very close to each other from the very beginning and seems to form continuous line.

- $\eta = 0.5$: It takes 521 iterations to converge, steps in the contour diagram are far in the beginning and then close to each other.

- $\eta = 0.9$: It takes 295 iterations to converge, steps in the contour diagram are far in the beginning and then close to each other.

- $\eta = 1.3$: It takes 205 iterations to converge, steps in the contour diagram seems to oscilate in the beginning and then follow straight line to minima.

- $\eta >= 2.1$: Program does not converge and blows out to infinity.

However, the error change is negligible in the cases when program converges.

Locally Weighted Linear Regression

(b) Analytical solution derived for locally weighted linear regression is,

$$\theta = (X^T W X)^{-1} X^T W Y$$

(c) $\tau = 0.8$ seems to work best. When τ is very small, it means that only points in the very near vicinity of the query point are considered. Where as, when τ is very large, the contribution of points far away from the query is also included which results in the generalization of the whole data means for very large τ values, it is almost same as the unweighted linear regression.

Logistic Regression

(a) The formula for the i^{th} row and j^{th} column element of the hessian matrix is,

$$\Sigma_{k=1 \text{ to } m} (-x_{[i]}^{(k)} * x_{[j]}^{(k)} * h_\theta(x^{(k)}) * (1 - h_\theta(x^{(k)})))$$

and, θ = [0.76037154, 1.17194674, -2.6205116]

Gaussian Discrmimant Analysis

(a)

- μ_0 : [137.46 , 366.62]
- μ_1 : [98.38 , 429.66]
- σ : [287.482 -26.748]
  [-26.748 1123.25]

(c) Equation of boundary :

$$2X^T \Sigma^{-1}(\mu_0 - \mu_1) = \mu_0^T \Sigma^{-1} \mu_0 - \mu_1^T \Sigma^{-1} \mu_1$$

(d)

- μ_0 : [137.46 , 366.62]
- μ_1 : [98.38 , 429.66]
- σ_0 : [319.5684 130.8348]
  [130.8348 875.3956]
- σ_1 : [255.3956 -184.3308]
  [-184.3308 1371.1044]

(e) Equation of boundary :

$$X^T (\Sigma_0^{-1} - \Sigma_1^{-1})x - 2X^T (\Sigma_0^{-1}\mu_0 - \Sigma_1^{-1}\mu_1) + \mu_0^T \Sigma_0^{-1} \mu_0 - \mu_1^T \Sigma_1^{-1} \mu_1 = 0$$

(f) On closely observing the graph, I find that the red points seem to cross the linear boundary. So, quadratic curve tries to include those points within its own region. That is why the curve is concave towards red points.

-------- END OF DOCUMENT---------