Assignment 1

Due 2014/10/08 23:59:59

Submit your work with the filename: "`ML_assignment-${number}_${your-student-id}_${your-name}`." Late submissions will not be accepted.

1. What is the difference in terms of the performance between the regression hypotheses based on the objective (1) $\arg_\theta \min \sum_{t=1}^{N} \left[r^{(t)} - h(\boldsymbol{x}^{(t)}; \theta) \right]^2$ and $\arg_\theta$ (2) $\min \sum_{t=1}^{N} \left| r^{(t)} - h(\boldsymbol{x}^{(t)}; \theta) \right|$ respectively?

2. In logistic regression, show that $l(\boldsymbol{\beta}) = \sum_{t=1}^{N} \left\{ y^{(t)} \boldsymbol{\beta}^\top \widetilde{\boldsymbol{x}}^{(t)} - \log \left(1 + e^{\boldsymbol{\beta}^\top \widetilde{\boldsymbol{x}}^{(t)}} \right) \right\}$.

3. Read Appendix C on the definitions of convex set and functions.

   (a) Show that the intersection of convex sets, $\bigcap_{i \in \mathbb{N}} C_i$ where $C_i \subseteq \mathbb{R}^n$, is convex.

   (b) Show that the log-likelihood function for logistic regression, $l(\boldsymbol{\beta})$, is concave.

4. Consider the locally weighted linear regression problem with the following objective:

$$\arg \min_{\boldsymbol{w} \in \mathbb{R}^{d+1}} \frac{1}{2} \sum_{i=1}^{N} l^{(i)} \left(\boldsymbol{w}^\top \begin{bmatrix} 1 \\ \boldsymbol{x}^{(i)} \end{bmatrix} - r^{(i)} \right)^2$$

$l^{(i)}$: locally weighted function
要predict的資料點與現有data的距離
距離越大，$l^{(i)}$值越小
距離為0時值為1

   local to a given instance $\boldsymbol{x}'$ whose label will be predicted, where $l^{(i)} = \exp(-\frac{(\boldsymbol{x}' - \boldsymbol{x}^{(i)})^2}{2\tau^2})$ for some constant τ.

   (a) Show that the above objective can be written as the form

$$(\boldsymbol{X}\boldsymbol{w} - \boldsymbol{r})^\top \boldsymbol{L} (\boldsymbol{X}\boldsymbol{w} - \boldsymbol{r}).$$

   Specify clearly what $\boldsymbol{X}$, $\boldsymbol{r}$, and $\boldsymbol{L}$ are.

   (b) Give a close form solution to $\boldsymbol{w}$. (Hint: recall that we have $\boldsymbol{w} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{r}$ in linear regression when $l^{(i)} = 1$ for all i)

   (c) Suppose that the training examples $(\boldsymbol{x}^{(i)}, r^{(i)})$ are i.i.d. samples drawn from some joint distribution with the marginal:

$$p(r^{(i)} | \boldsymbol{x}^{(i)}; \boldsymbol{w}) = \frac{1}{\sqrt{2\pi\sigma^{(i)}}} \exp \left(- \frac{\left(r^{(i)} - \boldsymbol{w}^\top \begin{bmatrix} 1 \\ \boldsymbol{x}^{(i)} \end{bmatrix} \right)^2}{2\sigma^{(i)2}} \right)$$

   where $\sigma^{(i)}$'s are constants. Show that finding the maximum likelihood of $\boldsymbol{w}$ reduces to solving the locally weighted linear regression problem above. Specify clearly what the $l^{(i)}$ is in terms of the $\sigma^{(i)}$'s.

where $\sigma^{(i)}$'s are constants. Show that finding the maximum likelihood of $\boldsymbol{w}$ reduces to solving the locally weighted linear regression problem above. Specify clearly what the $l^{(i)}$ is in terms of the $\sigma^{(i)}$'s.

(d) Implement a linear regressor (see the spec for more details) on the provided 1D dataset. Plot the data and your fitted line. (Hint: don't forget the intercept term)

(e) Implement 4 locally weighted linear regressors (see the spec for more details) on the same dataset with $\tau = 0.1$, 1, 10, and 100 respectively. Plot the data and your 4 fitted curves (for different $\boldsymbol{x}'$s within the dataset range).

(f) Discuss what happens when τ is too small or large.