Homework 1

Adheesh Juvekar, Arul Thileeban
September 18, 2019

Written Problems

1.

a. $P(y|x; W) = \frac{exp(w_y{}^T x)}{\sum_{c=1}^{C} exp(w_c{}^T x)}$

Calculate the likelihood function :

$L(D) \; = \; \prod_{i=1}^{n} \frac{exp(w_{y_i}{}^T x_i)}{\sum_{c=1}^{C} exp(w_c{}^T x_i)}$

To get log-likelihood function, take log on both sides.

$logL(D) \; = \; log(\prod_{i=1}^{n} \frac{exp(w_{y_i}{}^T x_i)}{\sum_{c=1}^{C} exp(w_c{}^T x_i)})$

$logL(D) \; = \; \sum_{i=1}^{n} log(\frac{exp(w_{y_i}{}^T x_i)}{\sum_{c=1}^{C} exp(w_c{}^T x_i)})$

$logL(D) \; = \; \sum_{i=1}^{n} (log(exp(w_{y_i}{}^T x_i)) - log(\sum_{c=1}^{C} exp(w_c{}^T x_i)))$

$logL(D) \; = \; \sum_{i=1}^{n} ((w_{y_i}{}^T x_i) - log(\sum_{c=1}^{C} exp(w_c{}^T x_i)))$

b. Differentiating the above function w.r.t w_c should give us the gradient of likelihood.

$\nabla_{w_c} L = \; \sum_{i=1}^{n} \; [y_i \; = \; \omega_c]x_i \; + \; \frac{(exp(w_C{}^T x_i)\,)* \, [y_i = \omega_c]x_i}{\sum_{c=1}^{C} exp(w_c{}^T x_i)}$

$= \; \sum_{i=1}^{n} \; [y_i \; = \; \omega_c]x_i \; + \; \frac{exp(w_C{}^T x_i)}{\sum_{c=1}^{C} exp(w_c{}^T x_i)} * [y_i \; = \; \omega_c]x_i$

$= \; \sum_{i=1}^{n} \; [y_i \; = \; \omega_c]x_i \; + [p(y = c|x_i; W)][y_i \; = \; \omega_c x_i$ (from (6) in question)

$= \sum_{i=1}^{n} x_i[y_i \; = \; \omega_c](1 \; + [p(y = c|x_i; W)])$

c. To avoid the overfitting effect, regularizers try to minimize the weight values and thus they ‘prefer’ the weight values to be zero so that the **probability of predicting each class is** $\frac{1}{C}$ **.**

Regularizers try to regularize the prediction, the regularizers try to reduce the phenomenon of overfitting, which is due to the model complexity (which is calculated by the weights of the model in this case).

This probability ($\frac{1}{C}$) is justified because, **it is better to assume the real data to be equally distributed among all the classes for generalization/regularization.**