EXAM 1 SOLUTIONS

Tuesday, October 7, 2014

Problem 1 (25 points)

$X$ is a scalar random variable with the following probability distribution,

$$p_{X|B}(x|b) = \begin{cases} bx^{-(b+1)} & x \geq 1 \\ 0 & x < 1 \end{cases},$$

where $b$ is another random variable, whose a priori distribution is

$$p_B(b) \propto \begin{cases} b^m \mu^{-(b+1)} & b > 0 \\ 0 & b \leq 0 \end{cases},$$

and where $\mu$ and $m$ are constant hyperparameters. You are given a database of i.i.d. training examples $X = \{x_1, \ldots, x_n\}$. The MAP estimate of $b$ is defined by $b_{MAP} = \arg \max p(b|X)$. Find $b_{MAP}$.

Solution

$$b_{MAP} = \frac{m + n}{m \ln \mu + \sum_{i=1}^{n} \ln x_i}$$

Problem 2 (25 points)

Discrete random variable $H$ and continuous random variable $V$ are jointly distributed as

$$p_{H,V|\Theta}(p, v|\theta) = \frac{0.5}{\sqrt{2\pi}} e^{-\frac{1}{2}(v-\mu_h)^2}, \quad h \in \{0, 1\}, \quad -\infty < v < \infty$$

where $\theta = [\mu_0, \mu_1]^T$ is a vector of parameters. You are given a database of i.i.d. examples of $V$, $V = \{v_1, \ldots, v_n\}$, but you are not told what are the associated values of $H$. Define

$$\gamma_i(h) = \frac{p_{H,V|\Theta}(h, v_i|\theta)}{p_{H,V|\Theta}(0, v_i|\theta) + p_{H,V|\Theta}(1, v_i|\theta)}$$

Suppose there are two candidate parameter vectors, $\theta$ and $\tilde{\theta} = [\tilde{\mu}_0, \tilde{\mu}_1]^T$, and suppose that

$$Q(\theta, \tilde{\theta}) = \sum_{i=1}^{n} E \left[ \ln p_{H,V|\tilde{\theta}}(H, v_i|\tilde{\theta}) | v_i, \tilde{\theta} \right]$$

Find $\partial Q/\partial \tilde{\mu}_0$. 


Solution

\[ \frac{\partial Q}{\partial \tilde{\mu}_0} = \sum_{i=1}^{n} \gamma_i(0)(v_i - \tilde{\mu}_0) \]

Problem 3  (25 points)

A two-layer neural net has MSE error criterion

\[ E = \frac{1}{2} \sum_{i=1}^{n} \|z_i - t_i\|^2 \]

where \( t_i = [t_{1i}, \ldots, t_{ri}]^T \) is the target vector, and \( z_i = [z_{1i}, \ldots, z_{ri}]^T \) is the network output. \( z_i \) is computed as

\[ z_{ki} = \text{g}_{\text{RLU}}(w_k^T y_i) \]

where \( w_k = [w_{1k}, \ldots, w_{qk}]^T \) is a weight vector, and \( y_i = [y_{1i}, \ldots, y_{qi}]^T \) is the hidden layer. \( y_i \) is computed as

\[ y_{ji} = \text{g}_{\text{RLU}}(v_j^T x_i) \]

where \( v_j = [v_{1j}, \ldots, v_{pj}]^T \) is a weight vector, and \( x_i = [x_{1i}, \ldots, x_{pi}]^T \) is the network input. The rectified linear units are defined by

\[ g_{\text{RLU}}(a) = \max(0, a) \]

Notice that, with these definitions,

\[ \frac{\partial E}{\partial w_{kj}} = \sum_{i \in S} \delta_{ki} y_{ji} \]

for some set of indices \( S \) which is a subset of \( \{1, \ldots, n\} \). Find a definition of \( S \) that permits you to write \( \delta_{ki} = (z_{ki} - t_{ki}) \).

Solution

\[ S = \{ i : w_k^T y_i > 0 \} \]

Problem 4  (25 points)

Second-order error approximations are defined by

\[ E(w) \approx \frac{1}{2}(w - w^*)^TB(w - w^*) + E_{\text{min}} \]

The line search algorithm is defined by

\[ \alpha_t = \arg \min_{\alpha} E(w_t + \alpha d_t) \]

\[ w_{t+1} = w_t + \alpha_t d_t \]
Let \( v_k \) and \( \lambda_k \) be the eigenvectors and eigenvalues, respectively, of the matrix \( B \), and define

\[
\begin{align*}
    r_{kt} &= v_k^T(w_t - w^*) \\
    q_{kt} &= v_k^T d_t
\end{align*}
\]

Express \( \alpha_t \) as a function of only \( r_{kt}, q_{kt} \), and \( \lambda_k \), with no other variables in your answer.

Solution

\[
\alpha_t = -\frac{\sum_{k=1}^{N} \lambda_k q_{kt} r_{kt}}{\sum_{k=1}^{N} \lambda_k q_{kt}^2}
\]

\[
\alpha_t = -\frac{\sum_{k=1}^{N} \lambda_k q_{kt} r_{kt}}{\sum_{k=1}^{N} \lambda_k q_{kt}^2}
\]