Final Exam

(1.) Non-negative data: Bayesian regression

Standard Bayesian regression does not impose any constraints on the sign of target variables $\vec{t}$. However, sometimes negative predictions are unacceptable, for example when we try to predict heights, weights, or concentrations. In this problem, we will develop a Bayesian regression approach in which target variables are constrained to be non-negative.

Consider

$$p(t|\vec{x},\vec{w},\beta) = A\,\mathcal{N}(t|\vec{w}^T\cdot\vec{\varphi}(\vec{x}),\beta^{-1}) \times \times\, \sigma(k_0\,\vec{w}^T\cdot\vec{\varphi}(\vec{x})) \quad (*)$$

, where

A is the normalization and $\sigma(a) = \frac{1}{1+e^{-a}}$ is the sigmoid function.

Note that in the $k_0 \to \infty$ limit, $\vec{w}^T\cdot\vec{\varphi}(\vec{x}) < 0$ are not allowed in (*).

(a) given $\{t_n\}_{n=1}^N$, where $t_n \geq 0$, $\forall n$, (all observations are non-negative)

compute the posterior probability

$p(\vec{w} \mid \underbrace{\vec{t}, X}_{\text{training data}}, \alpha, \beta)$, using

$p(\vec{w}) = \mathcal{N}(\vec{w} \mid \vec{0}, \alpha^{-1} \mathbb{I}_M)$ as the prior.

M: # weights & biases in the model

Use the Laplace approximation to ~~[illegible]~~ approximate the posterior with a Gaussian: $\mathcal{N}(\vec{w} \mid \vec{m}_N, S_N)$, and find $\vec{m}_N$ & S_N.

(b) Compute the predictive distribution, $p(t \mid \vec{x}, \underbrace{\vec{t}, X}_{\text{training data}}, \alpha, \beta)$ in the $k_0 \to \infty$ limit.

$\vec{x}$: new input pattern

② <u>Relevance vector machines for classification</u>

Show that in the $K=2$ case with binary target variables $t \in \{0,1\}$, the log marginal likelihood $\log p(\vec{t} \mid X, \vec{\alpha}) \equiv L(\vec{\alpha})$ can be written as $L(\vec{\alpha}) = L(\vec{\alpha}_{-i}) + \lambda(\alpha_i)$, where

$L(\vec{\alpha}_{-i})$ is the log marginal likelihood with basis function $\vec{\varphi}_i$ omitted, and $\lambda(\alpha_i)$ depends only on the hyperparameter α_i. Find $\lambda(\alpha_i)$ and use $\frac{d\lambda(\alpha_i)}{d\alpha_i}$ to

(i) analyze sparsity (i.e., discuss circumstances under which $\alpha_i \to \infty$ is the only solution and, conversely, when α_i may be finite).

(ii) Formulate a sequential sparse Bayesian learning algorithm for $K=2$ RVM classification.