Lab2

ECE 368

Bayesian Linear Regression

- Linear regression is the "work horse" of statistics and (supervised) machine learning.
- We assume the response is a linear function of the input, and we model our uncertainty with a Gaussian noise.

$$y(x) = \boldsymbol{a}^T x + w = \sum_{j=1}^{d} a_j x_j + w$$

Here, w is the additive Gaussian noise

$$w \sim \mathcal{N}(0, \sigma^2),$$

where σ^2 is a known parameter.

- We can also consider the linear model as a conditional probability distribution as

$$p(y|x, \boldsymbol{a}) = \mathcal{N}(\boldsymbol{a}^T x, \sigma^2)$$

- Suppose we are given a data set $\mathcal{D} = \{(x_1, y_1), \ldots, (x_N, y_N)\}$. Also, assume we have a prior over the weight vectors $\boldsymbol{a}$.
- **In this lab we want to use full Bayesian inference to make the prediction.**

Bayesian Linear Regression

The aim of Bayesian Linear Regression is not to find the single "best" value of the model parameters, but rather to determine the posterior distribution for the model parameters.
How to make prediction?

1. Computer the posterior

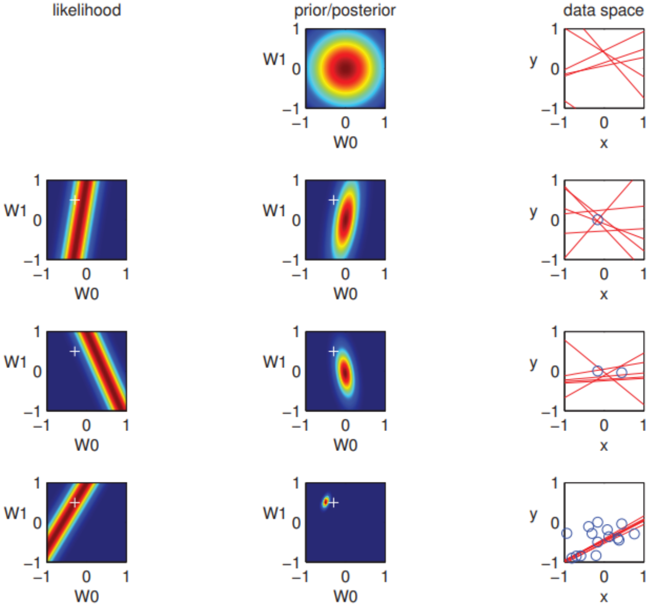
$$\underbrace{p(\boldsymbol{a}|\mathcal{D})}_{\text{Posterior}} \propto \underbrace{p(\boldsymbol{a})}_{\textit{Prior}} \underbrace{p(\mathcal{D}|\boldsymbol{a})}_{\textit{Likelihood}}$$

2. Make predictions using the posterior predictive distribution. Assume we are given a test point x, and we aim to predict its label y.

$$p(y|x, \mathcal{D}) = \int p(\boldsymbol{a}|\mathcal{D}) p(y|x, \boldsymbol{a}) d\boldsymbol{a}$$

Doing this lets us quantify our uncertainty

Visualization



Questions

- In this lab, we use Bayesian regression to fit a linear model. Consider a linear model of the form

$$z = a_1 x + a_0 + w, \tag{1}$$

where x is the scale input variable, and $\mathbf{a} = (a_0, a_1)^T$ is the vector-valued parameter with unknown entries a_0, a_1, and w is the additive Gaussian noise:

$$w \sim \mathcal{N}(0, \sigma^2), \tag{2}$$

where σ^2 is a known parameter.

- Suppose that we have access to a training data set containing N samples $\{x_1, z_1\}, \{x_2, z_2\}, \ldots, \{x_N, z_N\}$. We aim to estimate the parameter $\mathbf{a}$ by finding its posterior distribution. When the training finishes, we make predictions based on new inputs. We consider a Bayesian approach, which models the parameter $\mathbf{a}$ as a zero mean isotropic Gaussian random vector whose probability distribution is expressed as

$$p(\mathbf{a}) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \beta & 0 \\ 0 & \beta \end{bmatrix} \right), \tag{3}$$

where β is a known hyperparameter.

Questions

1. Express the posterior distribution $p(\mathbf{a}|x_1, z_1, \ldots, x_N, z_N)$ using σ^2, β, $x_1, z_1, x_2, z_2, \ldots, x_N, z_N$.
   Hint: We have $p(z_1, \ldots, z_N|\mathbf{a}, x_1, \ldots, x_N) = \Pi_{i=1}^N p(z_i|\mathbf{a}, x_i)$
   Hint: z is the sum of two independent Gaussian random variables.
2. Let $\sigma^2 = 0.1$ and $\beta = 1$. Based on the posterior distribution obtained in the last question, draw four contour plots corresponding to $p(\mathbf{a})$, $p(\mathbf{a}|x_1, z_1)$, $p(\mathbf{a}|x_1, z_1, \ldots, x_5, z_5)$, and $p(\mathbf{a}|x_1, z_1, \ldots, x_{100}, z_{100})$. In all contour plots, the x-axis represents a_0, and the y-axis represents a_1.
3. Suppose that there is a new input x, for which we want to predict the target value z. Write down the distribution of the prediction z, i.e., $p(z|x, x_1, z_1, \ldots, x_N, z_N)$.
4. Let $\sigma^2 = 0.1$ and $\beta = 1$. Suppose that the set of the new inputs is $\{-4, -3.8, -3.6, \ldots, 0, \ldots, 3.6, 3.8, 4\}$. Plot three figures corresponding to the following three cases:
   1. The predictions are based on one training sample, i.e., based on $p(z|x, x_1, z_1)$.
   2. The predictions are based on 5 training samples, i.e., based on $p(z|x, x_1, z_1, \ldots, x_5, z_5)$.
   3. The predictions are based on 100 training samples, i.e., based on $p(z|x, x_1, z_1, \ldots, x_{100}, z_{100})$.