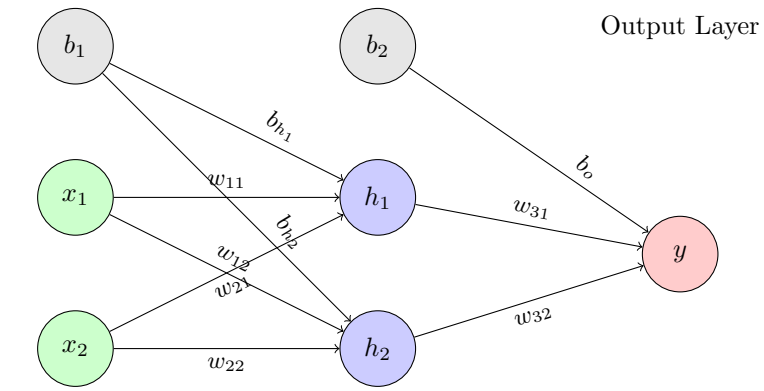
1 Linear Models and Neural Networks

1. **Ridge Regression:** Consider Ridge regression with a Gaussian prior on the parameter vector w and regularization parameter $\lambda \geq 0$. Given data points $(x_1, y_1), \ldots, (x_N, y_N)$, where each x_i is a D-dimensional vector:

   (a) Compute the exact solution for the parameter ω, use the loss function from the Maximum a Posteriori (MAP) value in Bayesian estimation

2. **Neural Network Analysis:** Consider the following neural network:



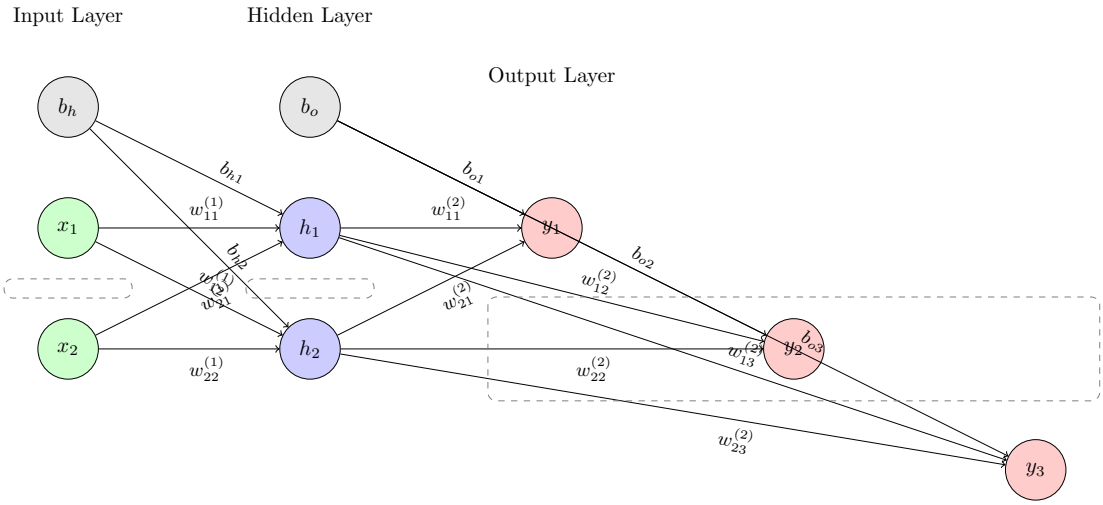
   (a) Compute the pre-activation (a) and activation (z) values for each hidden neuron and the output neuron

   (b) Given a data point $(\vec{x}, t)$, compute the mean squared error loss

   (c) Compute the error term (delta) for each neuron (MSE loss)

   (d) Compute the gradient for each parameter in the network

3. **Classification Neural Network:** Consider the following neural network for classification:



(a) Compute the pre-activation (a) and activation (z) values for each hidden neuron and the output neurons

(b) Compute the output probabilities using the softmax function

(c) Given a data point (x, t), compute the cross-entropy loss

(d) Using randomly generated data points (4 samples) and initial weights, with learning rate $\eta = 0.001$, compute the updated weights for five training steps

4. **Programming Assignment (PyTorch):** Implement and compare neural networks with logistic regression for classification on the CIFAR-10 dataset.

   Dataset: `https://docs.pytorch.org/vision/main/generated/torchvision.datasets.CIFAR10.html#torchvision.datasets.CIFAR10`

   Optimize the following hyperparameters to achieve the best test accuracy:

   - Network architecture (number of hidden layers, nodes per layer, activation functions)
   - Batch size
   - Number of training epochs
   - Learning rate

   Summarize your findings on how these parameters affect model performance.