# CS512 Machine Learning, Fall 2019
# Homework 3

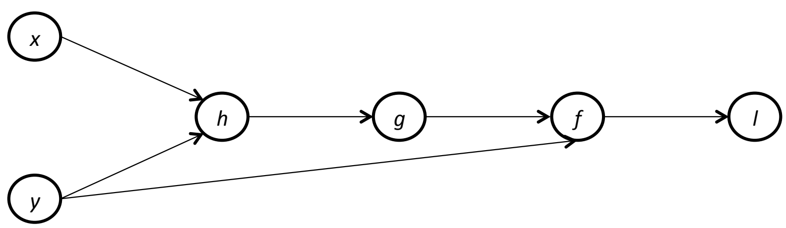
December 18, 2019

**Instructions**

- Submit an online copy of your report and code through SUCourse.
- Please submit the report and the code separately. Name your submission as CS512-YourName.pdf and CS512-YourName-Code.zip where you substitute in your first and last names into the filenames in place of 'YourName'.

## 1 Neural Networks [30 pts]

Consider the following neural network architecture for regression:




Here, $f(x,y)$ is the regression model, and $x$ and $y$ are two feature dimensions. $l(x,y)$ is the $\ell_2$ loss function. More specifically, the model is defined in terms of the following equations:

$$\begin{aligned} l(x,y) &= (f(x,y)-t)^2 \\ f(x,y) &= g(x,y)+yc \\ g(x,y) &= \sigma(h(x,y)) \\ h(x,y) &= ax+by \end{aligned}$$

where $t$ is the groundtruth label (i.e. target value), and $\sigma(s)$ is the sigmoid activation function:

$$\sigma(s) = \frac{1}{1+\exp(-s)}$$

**a) [20 pts]** As can be seen in the equations above, the model has three parameters: $a$, $b$, and $c$. Compute the partial derivatives of $l(x,y)$ with respect to the model parameters using back-propagation: $\frac{\partial l(x,y)}{\partial a}$, $\frac{\partial l(x,y)}{\partial b}$, and $\frac{\partial l(x,y)}{\partial c}$. Show your derivations step-by-step. (Hint: $\frac{\partial}{\partial s}\sigma(s) = \sigma(s)(1-\sigma(s))$)

**b) [10 pts]** Calculate the values of the loss and the partial derivatives with respect to the model parameters when $x=2$, $y=-1$, $a=1$, $b=2$, $c=-0.5$ and $t=0$ using back-propagation. Briefly show your forward and backward pass steps.

## 2 Spectral Clustering [70 pts]

**a)** **[40 pts]** Implement the spectral clustering algorithm function as follows:

| function[ `clusters`, `evals`, `evects` ] = `spectral(W, K, method)` | |
|---|---|
| **Parameters** | |
| $W$ | $N \times N$ similarity matrix |
| $K$ | the number of clusters |
| `method` | the method to be used for the Laplacian, 'normalized' |
| **Returned values** | |
| `clusters` | $1 \times N$ vector with cluster membership of each datapoint |
| `evals` | $1 \times K$ vector of smallest eigenvalues of the graph Laplacian |
| `evects` | $N \times K$ matrix of the corresponding eigenvectors of the graph Laplacian |
| **Decription** | |
| Clusters data points by the spectral clustering algorithm. | |
| `'unnormalized'` | Uses the graph Laplacian, $L = D - W$, where $D$ is the degree matrix $\sum_{i=1}^{N} W_{ij}$. |

Not required but you may extend this method with options for 'unnormalized', and 'symmetric' Laplacians later on.

**b)** **[20 pts]** Reproduce the Figure 1c (unnormalized Laplacian with Gaussian similarity function) in the Luxemburg Spectral clustering tutorial (available at SuCourse).