Double Cross Validation

```python
def doubleCrossValidation (D, k, H):
    allIdxs = np.arange(len(D))
    # Randomly split dataset into k outer folds
    outer_Idxs = np.random.permutation(allIdxs)
    outer_Idxs = outer_Idxs.reshape(k, -1)
    outer_acc = []
    for outer_fold in range(k):
        # Get all indexes for this outer fold
        outer_testIdxs = outer_Idxs[outer_fold,:]
        # Get all the other indexes for training
        outer_trainIdxs = np.array(set(allIdxs) - set(outer_testIdxs)).flatten()
        best_acc = 0
        best_H = -np.inf
        for hyperparameter in H:
            # Split the outer training data into k inner folds
            inner_Idxs = np.random.permutation(outer_trainIdxs)
            inner_Idxs = inner_Idxs.reshape(k, -1)
            inner_acc = []
            for inner_fold in range(k):
                # Get all indexes for this inner fold
                inner_testIdxs = inner_Idxs[inner_fold,:]
                # Get all the other indexes
                iTrainIdxs = np.array(set(outer_trainIdxs) - set(inner_testIdxs)).flatten()
                # Train the model on the inner training data
                model = trainModel(D[iTrainIdxs], hyperparameter)
                # Test the model on the inner testing data
                inner_acc.append(testModel(model, D[inner_testIdxs]))
            avg_inner_acc = np.mean(inner_acc)
            # The best hyperparameter will be selected from here
            if avg_inner_acc > best_acc:
                best_acc = avg_inner_acc
                best_H = hyperparameter
        # The best hyperparameter will be found from above and then implemented in the model training below
        model = trainModel(D[outer_trainIdxs], best_H)
        # Model tested on the outer model and then the accuracies are appended
        outer_acc.append(testModel(model, D[outer_testIdxs]))
    return np.mean(outer_acc)
```

Question 3 Code Output

```
 [-1.59158401]
 [-0.67883068]
 ...
 [ 1.47508802]
 [-0.6907794 ]
 [ 1.29762975]] alpha 0.001 mini_batch 39
the fmse value is :  144.77134166007122
```

Sumeet Shanbhag
ID:- 641020714

Deep Learning Proofs.

Q 2. a) $F(x,y) = x^4 + xy + x^2$
Prove that F is a ~~E~~ convex by proving its Hessian matrix is PSD or identify a point in the domain of F where Hessian is not PSD.

The Hessian matrix can be found out by :-

$$H = \begin{bmatrix} \partial^2 F / \partial x \partial x & \partial^2 F / \partial x \partial y \\ \partial^2 F / \partial y \partial x & \partial^2 F / \partial y \partial y \end{bmatrix}$$

$$\frac{\partial F}{\partial x} = 4x^3 + y + 2x \quad , \frac{\partial F}{\partial x \partial x} = 12x^2 + 2$$

$$\frac{\partial F}{\partial y} = x \quad , \frac{\partial F}{\partial y \partial y} = 1.$$

$$\therefore H = \begin{bmatrix} 12x^2 + 2 & 1 \\ 1 & 0 \end{bmatrix}$$

We know that the Hessian ~~of~~ matrix determinant has to be positive or 0 for it to be positive semi definite.

$$\Delta_H = (12x^2 + 2) \times 0 - 1 \times 1 = -1$$

Since the determinant of the matrix is -1, this means that there exists a point where the Hessian is negative. Thus F is not a convex function.

2b) Show that, for any real vector v, it is always true that $v^T H v \geq 0$.

$$F_{MSE}(w) = \frac{1}{2n}(X^T w - y)(X^T w - y)$$

$$= \frac{1}{2n}((X^T w - y))^2$$

$$\frac{\partial F_{MSE}(w)}{\partial w} = \frac{1}{\cancel{2}n} \times \cancel{2} \times (X) \times (X^T w - y)$$

$$\frac{\partial^2 F_{MSE}(w)}{\partial w^2} = \frac{1}{\cancel{2}n} \times (X) \times (X^T) \quad = \text{Hessian matrix.} = H$$

To prove that $v^T H v \geq 0$,

$$v^T H v = (1/n) \times (v^T) \times (X) \times (X^T) \times v$$

$X . X^T$ is a positive semi definite matrix which means that $(v^T) * (X) * (X^T) * (v)$ will also be ≥ 0.

This means that H is also PSD, which means the func^n $F_{MSE}(w)$ is convex wrt weight vector w.

4a) $\sigma(x) = \frac{1}{1+e^{-x}}$

Prove that $\sigma(-x) = 1 - \sigma(x) \quad \forall x.$

$$LHS = \sigma(-x) = \frac{1}{1+e^{x}}$$

$$RHS = 1 - \sigma(x) = 1 - \frac{1}{1+e^{-x}}$$

$$= \frac{1+e^{-x}-1}{1+e^{-x}}$$

$$= \frac{e^{-x}}{1+e^{-x}}$$

Multiplying by $\frac{e^{x}}{e^{x}}$, we get

$$= \frac{e^{-x} \times e^{x}}{(1+e^{-x})e^{x}}$$

$$= \frac{1}{1+e^{x}}$$

$\therefore LHS = RHS$

4b) Prove that $\sigma'(x) = \frac{\partial \sigma(x)}{\partial x} = \sigma(x)(1-\sigma(x)) \quad \forall x$

$$\frac{\partial \sigma(x)}{\partial x} = \frac{1}{(1+e^{-x})^2} \times e^{-x} \times (-1)$$

$$= \frac{-e^{-x}}{(1+e^{-x})^2} = \text{LHS}.$$

$$\text{RHS} = \sigma(x)(1-\sigma(x))$$

$$= \frac{1}{(1+e^{-x})}\left(1 - \frac{1}{(1+e^{-x})}\right)$$

$$= \frac{1}{(1+e^{-x})}\left(\frac{1+e^{-x}-1}{1+e^{-x}}\right)$$

$$= \frac{e^{-x}}{(1+e^{-x})^2}$$

$\therefore$ LHS = RHS