**CECS 551**
**Assignment 5**
**Total: 20 Points**

---

General Instruction

- Submit uncompressed file(s) in the Dropbox folder via BeachBoard (Not email).

---

1. (20 points) Find best hyper-parameters using `GridSearchCV`.

   - Find `Assignment_5.ipynb`.
   - The toy dataset and the network design are identical with the assignment 3.
   - Adam optimizer is used for the optimization, and you need to determine three parameters, `beta_1, beta_2, learning_rate`.
   - Adam optimizer is given in Figure 1. `beta_1` corresponds to $\rho_1$, `beta_2` corresponds to $\rho_2$, `learning_rate` corresponds to $\epsilon$.
   - You are asked to complete the last part of `Assignment_5.ipynb` to perform a grid search using `GridSearchCV`.
   - You need to have at least three values for each hyper-parameters.
   - Print out the 'negative mean squared error' and the corresponding hyper-parameters.

```
-0.046273 {'beta_1': 0.9, 'beta_2': 0.999, 'learning_rate': 0.0001}
-0.003971 {'beta_1': 0.9, 'beta_2': 0.999, 'learning_rate': 0.001}
-0.041319 {'beta_1': 0.9, 'beta_2': 0.999, 'learning_rate': 0.01}
    .
    .
-0.003831 {'beta_1': 0.8, 'beta_2': 0.99, 'learning_rate': 0.01}
    .
    .
Best: -0.003831 {'beta_1': 0.8, 'beta_2': 0.99, 'learning_rate': 0.01}
```

$t \leftarrow t + 1$
Update biased first moment estimate: $\boldsymbol{s} \leftarrow \rho_1 \boldsymbol{s} + (1 - \rho_1)\boldsymbol{g}$
Update biased second moment estimate: $\boldsymbol{r} \leftarrow \rho_2 \boldsymbol{r} + (1 - \rho_2)\boldsymbol{g} \odot \boldsymbol{g}$
Correct bias in first moment: $\hat{\boldsymbol{s}} \leftarrow \frac{\boldsymbol{s}}{1-\rho_1^t}$
Correct bias in second moment: $\hat{\boldsymbol{r}} \leftarrow \frac{\boldsymbol{r}}{1-\rho_2^t}$
Compute update: $\Delta\boldsymbol{\theta} = -\epsilon\frac{\hat{\boldsymbol{s}}}{\sqrt{\hat{\boldsymbol{r}}}+\delta}$ (operations applied element-wise)
Apply update: $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \Delta\boldsymbol{\theta}$

Figure 1: Adam optimizer