CSCI 362: Machine Learning **Assignment 1**
Fall 2023 **Sampling distributions**

Please submit a single document for this assignment on Canvas by the beginning of class on Wednesday, Sept. 6th.

The torch function `randn` can be used to generate normally distributed numbers.

1. Describe in one or two sentences the output of following program.

```
import torch

xs = torch.randn(30)
print(xs)
```

   Why is the output different each time you run the program?

2. Add lines to the program above that print the mean and standard deviation of the 30 numbers held in `xs`. Hint: PyTorch tensors implement methods `mean()` and `std()`. Include a screenshot of your program and its output in your document for this assignment.

   - Is the mean exactly equal to zero? If it is not, why not.
   - Is the standard deviation exactly equal to one? If it is not, why not.

   Write your answers in your document.

3. Modify your program so that it generates and prints 30 numbers from the normal distribution with mean 100 and standard deviation 25. (Note: we saw in class that there is more than one way to do this in PyTorch!) Include a screenshot of your program and its output in your document for this assignment.

4. If you use your program in the last exercise to generate a single *sample* of size 30, it is highly unlikely that the mean of the sampled numbers will be exactly 100. In fact, you would routinely get numbers in the low 90s or around 110.

   However, suppose that you draw many samples of size 30; and that for each such sample you compute and record its mean. What would you expect the *mean of those means* to be?

   Write a program do this and see if your prediction is correct? (Screenshot your code and its output and include it in your document).

5. Do the same as in the last exercise but for the standard deviation. Does the *mean of the standard deviations* over very many samples target the correct value? As always, include your code and its output.

6. What happens if you sample from the uniform distribution on [0,1] instead of a normal distribution? What value does the mean of the means of samples of size 30 appear to target? What about the mean of the standard deviations of many samples of size 30?

   Modify your program above, and include screenshots in your solution document.