HW #4

1. (a) Consider Bessel functions of the 1^{st} kind:

$$J_\alpha(x) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!\,\Gamma(m+\alpha+1)} \left(\frac{x}{2}\right)^{2m+\alpha}.$$

Use $y(x) = J_0(x) + \epsilon$ to generate $N = 500$ datapoints in the $(0, 30]$ range, with x evenly spaced within the range. Here, $\epsilon \sim \mathcal{N}(0, \sigma^2)$, with $\sigma = 0.05$.

Plot your dataset superimposed on the true function $J_0(x)$.

(b) Use the dataset from (a) to fit a GP model with:

(i) RBF kernel: $k(x_n, x_m) = e^{-\frac{\theta}{2}(x_n - x_m)^2}$

(ii) exponential kernel: $k(x_n, x_m) = e^{-\theta|x_n - x_m|}$

→ Find θ^*, the optimal value of θ, for both kernels by maximizing $\log p(\tilde{y}|\theta)$ wrt θ. Report θ^* in both cases.

→ Using $\theta = \theta^*$, find mean $(\tilde{\mu})$ and std_dev $(\tilde{\sigma})$ of the predictive distribution for both kernels.

Plot $J_0(x)$, $\tilde{y} = \overbrace{y_1, \ldots, y_N}$, and

$\tilde{\mu} \pm \tilde{\sigma}$ curves separately for each kernel, in the $[-10, 40]$ range.

(2.) Fit the dataset from 1a using an MLP with a single hidden layer. (Using existing packages is allowed)

→ Use $M = 8, 16, 32, 64$ (M = # hidden nodes)

→ Split the data into training ($N_{tr} = 400$) and test ($N_{ts} = 100$) subsets.

(a) → For each NN architecture, train the model 16 times starting from random weights and biases (use <u>Xaveier initialization</u>)

→ plot the average sum-of-squares error on the test subset vs. M.

→ Use early stopping or weight decay to regularize the models. Use an activation function of your choice; choose any optimizer.

(b) Repeat part (a), but with another activation function (specify which!)

(c) Repeat part (a), but with another optimizer (specify which!)