00137960: Statistical Thinking
Homework 1

1. The *general significant-digit law* says that for all $k \in \mathbb{N}$, $d_1 \in \{1, \dots, 9\}$, $d_j \in \{0, 1, \dots, 9\}$, $j = 2, \dots, k$,
$$P\left(\bigcap_{j=1}^{k} \{D_j = d_j\}\right) = \log_{10}\left\{1 + \left(\sum_{j=1}^{k} d_j \cdot 10^{k-j}\right)^{-1}\right\},$$
where D_j is the jth significant digit. Use this result to deduce that the distribution of D_j approaches the uniform distribution as $j \to \infty$.

2. Discuss the identifiability of the following models and whether the nonidentifiability can be eliminated by some simple assumptions that restrict the parameter space:
   (a) the additive model
   $$y = \beta_0 + \sum_{j=1}^{p} f_j(x_j) + \varepsilon,$$
   where $\beta_0 \in \mathbb{R}$ and $f_j : \mathbb{R} \to \mathbb{R}$ are unknown functions;
   (b) the single-index model
   $$y = g(x^T \beta) + \varepsilon,$$
   where $\beta \in \mathbb{R}^p$ and $g : \mathbb{R} \to \mathbb{R}$ is an unknown function;
   (c) the two-layer neural network
   $$y = \sum_{j=1}^{m} a_j \sigma(x^T w_j + b_j) + c,$$
   where $a_j, b_j, c \in \mathbb{R}$, $w_j \in \mathbb{R}^p$, and $\sigma(z) = \max(z, 0)$ is the ReLU activation function.

3. The temperature T at location s is modeled as a stationary isotropic Gaussian process with mean $ET(s) = \mu$ and covariance function $\mathrm{Cov}(T(s_1), T(s_2)) = \sigma^2 \exp(-\gamma |s_1 - s_2|)$, where the parameter space is $(\mu, \sigma^2, \gamma) \in \mathbb{R} \times (0, \infty)^2$ and inference for $\theta = \mu/\sigma$ is required. Explain, through McCullagh's formal theory of statistical models, why this model is not sensible.

4. Suppose we fit a linear model $y = x^T \beta + \varepsilon$ to data generated from $y = f(x) + \varepsilon$, where $E\varepsilon = 0$ and $\mathrm{Var}(\varepsilon) = \sigma^2$. Verify that the prediction error at $x = x_0$ can be decomposed as
$$\text{Prediction Error} = (\text{Model Bias})^2 + (\text{Estimation Bias})^2 + \text{Variance} + \sigma^2.$$

5. (Programming) Reproduce the experiments in Biecek et al. (2024, "Performance Is Not Enough: The Story Told by a Rashomon Quartet," *JCGS*, 33(3), 1118–1121): simulate data from
$$y = \sin\left(\frac{3x_1 + x_2}{5}\right) + \varepsilon,$$
where $\varepsilon \sim N(0, 1/3)$, $(x_1, x_2, x_3) \sim N_3(0, \Sigma)$, and
$$\Sigma = \begin{pmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \end{pmatrix};$$
fit each of the four models (linear model, decision tree, random forest, and neural network) to the data; report the predictive performance and draw partial dependence plots similar to Figure 2; and explain the *Rashomon effect*.