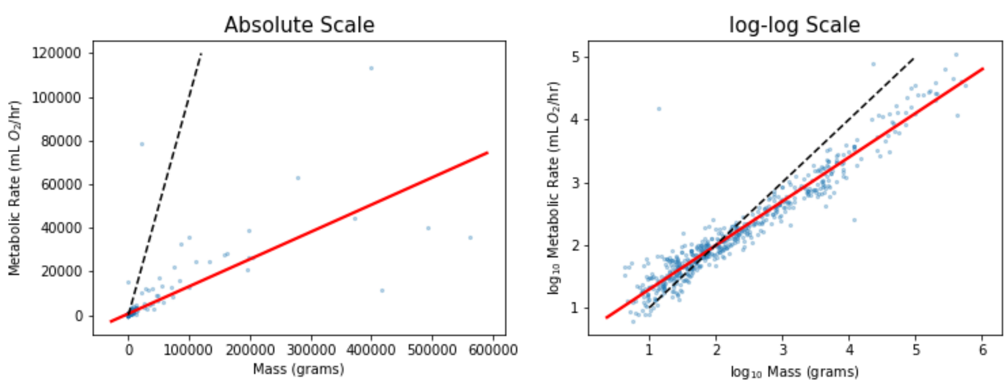
Data 100, Spring 2022	February 18, 2022

Discussion #5

Logarithmic Transformations

1. One of your friends at a biology lab asks you to help them analyze `panTHERIA`, a database of mammals. They are interested in the relationship between mass, measured in grams, and metabolic rate ("energy expenditure"), measured by oxygen use per hour. Originally, they show you the data on a linear (absolute) scale, shown on the left. You notice that the values on both axes vary over a large range with many data points clustered around the smaller values, so you suggest that they instead plot the data on a log-log scale, shown on the right. The solid red line is a "line of best fit" (we'll formalize this later in the course) while the black dashed line represents the identity line $y = x$.



   (a) Let C and k be some constants and x and y represent mass and metabolic rate, respectively. Based on the plots, which of the following best describe the pattern seen in the data? Reminder: $log(ab) = log(a) + log(b)$.

   ◯ A. $y = C + kx$ ◯ B. $y = C \times 10^{kx}$ ◯ C. $y = C + k\log_{10}(x)$ ◯ D. $y = Cx^k$

   (b) What parts of the plots could you use to make initial guesses on C and k?

   (c) Your friend points to the solid line on the log-log plot and says "since this line is going up and to the right, we can say that, in general, the bigger a mammal is, the greater its metabolic rate". Is this a reasonable interpretation of the plot?

Kernel Density Estimation

2. We wish to compare the results of kernel density estimation using a Gaussian kernel and a boxcar kernel. For $\alpha > 0$, which of the following statements are true? Choose all that apply.

   Gaussian Kernel:
   $$K_\alpha(x, z) = \frac{1}{\sqrt{2\pi\alpha^2}} \exp\left(-\frac{(x-z)^2}{2\alpha^2}\right)$$

   Boxcar Kernel:
   $$B_\alpha(x, z) = \begin{cases} \frac{1}{\alpha} & \text{if } -\frac{\alpha}{2} \leq x - z \leq \frac{\alpha}{2} \\ 0 & \text{else} \end{cases}$$

   A. Decreasing α for a Gaussian kernel decreases the smoothness of the KDE.

   B. The Gaussian kernel is always better than the boxcar kernel for KDEs.

   C. Because the Gaussian kernel is smooth, we can safely use large α values for kernel density estimation without worrying about the actual distribution of data.

   D. The area under the boxcar kernel is 1, regardless of the value of α.

   E. None of the above.

Driving with a Constant Model

3. Adam is trying to use modeling to drive his car autonomously. To do this, he collects a lot of data where he drives around his neighborhood, and he wants your help to design a model that can drive on his behalf in the future using the outputs of the models you design. We will tackle two aspects of this autonomous car modeling framework: going forward and turning.

   We show some statistics from the collected dataset below using `pd.describe`, which returns the mean, standard deviation, quartiles, minimum and maximum for the two columns in the dataset: `target_speed` and `degree_turn`.

	target_speed	**degree_turn**
count	500.000000	500.000000
mean	32.923408	143.721153
std	46.678744	153.641504
min	0.231601	0.000000
25%	12.350025	6.916210
50%	25.820689	45.490086
75%	39.788716	323.197168
max	379.919965	359.430309

   (a) Suppose the first part of the model predicts the target speed of the car. Using constant models trained on the speeds on the collected data shown above with L_1 and L_2 loss functions, which of the following is true?

   ◯ A. The model trained with the L_1 loss will always drive slower than the model trained with L_2 loss.

   ◯ B. The model trained with the L_2 loss will always drive slower than the model trained with L_1 loss.

   ◯ C. The model trained with the L_1 loss will sometimes drive slower than the model trained with L_2 loss.

   ◯ D. The model trained with the L_2 loss will sometimes drive slower than the model trained with L_1 loss.

   (b) Finding that the model trained with the L_2 loss drives too slowly, Adam changes the loss function for the constant model where the loss is penalized **more** if the speed is higher.

That way, the model wants to optimize more for the case where we wish to drive faster since the loss is higher, accomplishing his goal.

Find the optimal $\hat{\theta}$ for the constant model using the new loss function below:

$$L(\theta) = \frac{1}{n} \sum_i y_i (y_i - \theta)^2$$

(c) Suppose he is working on a model that predicts the degree of turning at a particular time between 0 and 359 degrees using the data in the `degree_turn` column. Explain why a constant model is likely inappropriate in this use case.

Extra: If you've studied some physics, you may recognize the behaviour of our constant model!

(d) Suppose we finally expand our modeling framework to use simple linear regression (i.e. $f_\theta(x) = \theta_{w,0} + \theta_{w,1} x$). For our first simple linear regression model, we predict the turn angle (y) using target speed (x). Our optimal parameters are: $\hat{\theta}_{w,1} = 0.019$ and $\hat{\theta}_{w,0} = 143.1$.

However, we realize that we actually want a model that predicts target speed (our new y) using turn angle, our new x (instead of the other way around)! What are our new optimal parameters for this new model?

Standardized SLR (Bonus)

4. We will experiment with standardizing the x (explanatory) and y (response) variables for a simple linear model with and without an intercept term (i.e. $f_\theta(x) = \theta x$). Recall that standardizing a variable V involves subtracting its mean $\bar{V}$ and dividing by its standard deviation σ_V as follows: $\tilde{V} = \frac{V - \bar{V}}{\sigma_V}$.

   (a) What is the optimal solution for θ with $f_\theta(x) = \theta x$ with a standard MSE loss for standardized x and y?

   (b) Calculate the optimal θ_0 and θ_1 for an SLR model with an intercept using if x and y were standardized.

   (c) Show that the optimal θ for a linear model without an intercept (i.e. $f_\theta(x) = \theta x$) is the same as the optimal θ for standardized SLR from the previous subpart.

   Hint: Given that the variance of the values in a vector $\vec{v}$ is $\sigma_v^2 = \frac{1}{n}\sum_i (v_i - \bar{v})^2$, simplify the denominator of the optimal θ from part (a).