Homework 7

Due Monday, March 7 before 5:00pm

Use [Live Editor > Save > Export to PDF] to prepare your submission for Gradescope.

This assignment uses data from the MAT file HW7_data.`mat`. Download this file and run

```
load HW7_data.mat
```

to load variable `blood` into the workspace.

Blood Metabolite Diagnostic for Fungal Infections

You are asked to design a diganostic for bloodborne fungal infections. Ideally, you would measure the number of colony forming units (CFUs) of fungus per ml of blood. However, the fungus is slow growing outside the body, so accurate CFU counts take weeks. Instead, you hope to use standard measurements from a blood metabolic panel to predict the CFUs/ml in a sample.

The Matlab table blood contains data from a 250-patient clinical trial. Each datapoint has values for all 14 standard blood metabolite readings:

Metabolite	Variable Name	Units
albumin	albumin	g/dL
alkaline phosphatase	alk_phos	IU/L
alanine aminotransferase	ALT	IU/L
aspartate aminotransferase	AST	IU/L
blood urea nitrogen	BUN	mg/dL
calcium	Ca	mg/dL
chloride	Cl	mmol/L
carbon dioxide	CO_2	mmol/L
creatinine	creatinine	mg/dL
glucose	glucose	mg/dL
potassium	K	mEq/L
sodium	Na	mEq/L
total bilirubin	bilirubin	mg/dL
total protein	protein	g/dL

The blood table also contains the log(CFU) counts for each sample. (*Note that we use log(CFU) since CFU counts vary exponentially.*)

a.) Using linear regression, build a model that predicts log(CFU) counts with blood metabolite readings. You do not need to include interactions in your model.

```
% place your code here
```

For this example, we will assess the statistical significance of the coefficients by considering only those with $p<0.05$. Which metabolite readings are significantly predictive of the CFU counts? Do these metabolite levels increase or decrease as the fungus count increases?

<place your answer here>

b.) What is the RMSE for your model? What are its units?

<place your answer here>

c.) Build another model using only the significant predictors. Does the RMSE change?

<place your answer here>

d.) During sepsis, the number of fungal cells in the blood increases by 100 fold. Would your original model be able to predict this level of change using metabolites? Why or why not?

<place your answer here>