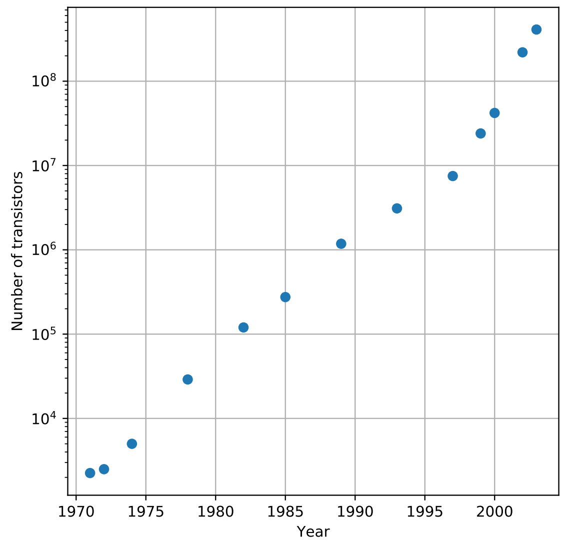
Homework 1

1. *Moore's law.* The figure and table below show the number of transistors N in 13 microprocessors, and the year of their introduction.

Year	# of transistors
1971	2,250
1972	2,500
1974	5,000
1978	29,000
1982	120,000
1985	275,000
1989	1,180,000
1993	3,100,000
1997	7,500,000
1999	24,000,000
2000	42,000,000
2002	220,000,000
2003	410,000,000



The plot gives the number of transistors on a logarithmic scale. Find the least squares straight-line fit of the data using the model

$$\log_{10} N \approx \theta_1 + \theta_2(t - 1970)$$

where t is the year and N is the number of transistors. Note that θ_1 is the model's prediction of the log of the number of transistors in 1970, and 10^{θ_2} gives the model's prediction of the fractional increase in number of transistors per year.

(a) Find the coefficients θ_1 and θ_2 that minimize the RMS error on the data, and give the RMS error on the data. Plot the model you find along with the data points.

(b) Use your model to predict the number of transistors in a microprocessor introduced in 2015. Compare the prediction to the IBM Z13 microprocessor, released in 2015, which has around 4×10^9 transistors.

(c) Compare your result with Moore's law, which states that the number of transistors per integrated circuit roughly doubles every one and a half to two years.

The computer scientist and Intel corporation co-founder Gordon Moore formulated the law that bears his name in a magazine article published in 1965.

2. *Nonlinear auto-regressive model.* We have a scalar time series $z_1, z_2, \ldots, z_T$. The following one step ahead prediction model is proposed:

$$\hat{z}_{t+1} = \theta_1 z_t + \theta_2 z_{t-1} + \theta_3 z_t z_{t-1}$$

where $\theta = (\theta_1, \theta_2, \theta_3)$ is the model parameter vector. The sum of the squares of the prediction error of this model on the given time series is

$$\sum_{t=2}^{T-1} (\hat{z}_{t+1} - z_{t+1})^2$$

Note that we must start this sum with $t = 2$, since z_0 and z_{-1} are not defined. Express this quantity as $\|A\theta - b\|_2^2$, where A is a $(T-2) \times 3$ matrix and b is a $(T-2)$-vector. (You must say what the entries of A and b are. They can involve the known data $z_1, \ldots, z_T$.)

3. *Minimizing a squared norm plus an affine function.* Minimizing a squared norm plus an affine function. A generalization of the least squares problem adds an affine function to the least squares objective,

$$\text{minimize} \quad \|Ax - b\|^2 + c^T x + d$$

where the n-vector x is the variable to be chosen, and the (given) data are the $m \times n$ matrix A, the m-vector b, the n-vector c, and the number d. We will use the same assumption we use in least squares: The columns of A are linearly independent. This generalized problem can be solved by reducing it to a standard least squares problem, using a trick called completing the square. Show that the objective of the problem above can be expressed in the form

$$\|Ax - b\|^2 + c^T x + d = \|Ax - b + f\|^2 + g$$

for some m-vector f and some constant g. It follows that we can solve the generalized least squares problem by minimizing $\|Ax - (b - f)\|$, an ordinary least squares problem with solution $\hat{x} = A^\dagger (b - f)$.

Hints. Express the norm squared term on the right-hand side as $\|(Ax - b) + f\|^2$ and expand it. Then argue that the equality above holds provided $2A^T f = c$. One possible choice is $f = (1/2)(A^\dagger)^T c$. (You must justify these statements.)

4. *Sequential outlier removal.* Throughout this problem, you'll use the data `U`, `v`, found in `fitting_outliers.json`. Here, $U \in \mathbf{R}^{n \times 1}$, so there is only one (nonconstant) feature. This one feature is already (nearly) standardized, so you do not need to standardize it. The data matrix X will have two columns, the constant feature one and the feature given in U. Also, there is enough data that you do not need to use any regularization.

(a) Fit a least squares model to the dataset above and plot the data points and straight-line fit. Describe what you observe.

(b) *Sequential outlier removal.* Find the data point with the largest loss and label it as an outlier. Remove this point from your data set and fit the model again to this new dataset (which has one fewer data point). Continue doing this until your θ stops changing too much (say, the change between the components of the previous θ and the current one is no more than .01).

Show a few of the intermediate fits and the final fit plotted against the data points. Describe what you observe.