Spring 2022

Stanford
AA 203: Optimal and Learning-based Control
Problem set 0

Problem 1: Poisson Maximum Likelihood

Suppose we observe the number of customers to a store over n days $x_1, x_2, ..., x_n$, and we want to fit a Poisson distribution to this data. The Poisson distribution is a distribution over non-negative integers with a single parameter $\lambda \geq 0$. It is often used to model arrival times of random events or count the number of random arrivals within a given amount of time. It has probability mass function:

$$\mathbb{P}_\lambda[X = k] = \frac{e^{-\lambda}\lambda^k}{k!} \text{ when } X \sim \text{Poi}(\lambda).$$

One way to do this is via *Maximum Likelihood*, where we choose the parameter of the Poisson distribution to maximize the probability that the data $x_1, x_2, ..., x_n$ appears. This can be done by maximizing the log-likelihood of the dataset $x_1, ..., x_n$ with respect to λ. The log-likelihood of $x_1, ..., x_n$ under the Poisson model is

$$f_\lambda(x_1, ...x_n) := \sum_{i=1}^{n} \ln\left(\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right).$$

Compute the Maximum Likelihood estimator $\hat{\lambda}$ by finding a solution to

$$\arg\max_{\lambda \geq 0} f_\lambda(x_1, ..., x_n).$$

Solution: First observe that

$$f_\lambda(x_1, ..., x_n) = -\lambda n + \ln(\lambda)\left(\sum_{i=1}^{n} x_i\right) - \sum_{i=1}^{n} \ln(x_i!)$$

is a concave function of λ. Therefore to find the maximum, we just need to find where $\frac{df_\lambda}{d\lambda}$ equals zero.

$$\begin{aligned} \frac{df_\lambda}{d\lambda} &= 0 \\ \iff -n + \frac{1}{\lambda}\left(\sum_{i=1}^{n} x_i\right) &= 0 \\ \iff \lambda = \frac{1}{n}\sum_{i=1}^{n} x_i&. \end{aligned}$$

Problem 2: Discrete Linear Systems

Consider the discrete linear system $x_{t+1} = Ax_t + Bu_t$, where

$$A = \begin{bmatrix} \frac{4}{5} & 0 & 0 \\ 0 & \sqrt{3} & 1 \\ 0 & -1 & \sqrt{3} \end{bmatrix}, B = \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

a) In the absence of control (i.e. $u_t = 0$), is this system stable? Why or why not?

Solution: The system is not stable in the absence of control. The eigenvalues of A are $4/5, 2e^{i\pi/6}, 2e^{-i\pi/6}$. An autonomous discrete system is stable if and only if the modulus of all eigenvalues of A are less than 1. Since $|2e^{i\pi/6}| = 2$ and $|2e^{-i\pi/6}| = 2$, this system is not stable.

b) Design a linear feedback controller $u_t = Kx_t$ for some fixed matrix $K \in \mathbb{R}^{2\times 3}$ so that the closed loop system will be stable.

Solution: The closed loop dynamics will be $x_{t+1} = Ax_t + BKx_t = (A + BK)x_t$. Recall from part 2.1 that the eigenvalues of A are $4/5, 2e^{i\pi/6}, 2e^{-i\pi/6}$. We need the eigenvalues of $A + BK$ to have modulus less than one, so we proceed by using BK to cancel the components of A corresponding to the eigenvalues $2e^{i\pi/6}, 2e^{-i\pi/6}$. First, note that

$$\begin{bmatrix} 0 & 1 \\ 1 & \text{-}1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{-1}.$$

Using this, let

$$K' := -\begin{bmatrix} 0 & 1 \\ 1 & \text{-}1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 1 \\ -1 & \sqrt{3} \end{bmatrix} \text{ and } K := \begin{bmatrix} 0 & K'_{11} & K'_{12} \\ 0 & K'_{21} & K'_{22} \end{bmatrix}.$$

By this construction, the bottom right submatrix of BK will be:

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} K' = -\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & \text{-}1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 1 \\ -1 & \sqrt{3} \end{bmatrix} = -\begin{bmatrix} \sqrt{3} & 1 \\ -1 & \sqrt{3} \end{bmatrix}$$

Thus we see that $A + BK$ has eigenvalues $4/5, 0, 0$ and is thus stable.

$$A + BK = \begin{bmatrix} \frac{4}{5} & 0 & 0 \\ 0 & \sqrt{3} & 1 \\ 0 & -1 & \sqrt{3} \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sqrt{3} & 1 \\ 0 & -1 & \sqrt{3} \end{bmatrix} = \begin{bmatrix} \frac{4}{5} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Problem 3: Linear Regression

Recall that the least squares solution to $\min_x \|Ax-b\|_2^2$ is given by the normal equation $x^* = (A^\top A)^{-1}A^\top b$.

a) Suppose in addition to finding an x so that Ax is close to b, we prefer x to be "small" as measured by $x^\top \Lambda x$, where Λ is a positive definite matrix. This gives rise to the *ridge regression* problem:

$$\min_x \|Ax-b\|_2^2 + x^\top \Lambda x.$$

Derive the normal equation (i.e. closed form solution) for the ridge regression problem.

Solution: Since Λ is positive definite, the objective function is convex. Therefore we can find the optimal solution by finding the point where the gradient is zero. Letting $f(x) := \|Ax-b\|_2^2 + x^\top \Lambda x$, we have

$$\begin{aligned}
\nabla_x f(x) &= \nabla_x \left(x^\top A^\top A x - 2x^\top A^\top b + b^\top b + x^\top \Lambda x \right) \\
&= \nabla_x \left(x^\top \left(A^\top A + \Lambda \right) x - 2x^\top A^\top b + b^\top b \right) \\
&= 2\left(A^\top A + \Lambda \right) x - 2A^\top b
\end{aligned}$$

Hence $\nabla_x f(x) = 0$ when $x = \left(A^\top A + \Lambda\right)^{-1} A^\top b$. Indeed, when there is no regularization term (i.e. $\Lambda = 0$), the ridge regression solution reduces to the normal equation of least squares.

b) We obtain measurements $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ on n asteroids, where x_i, y_i are estimates of the diameter and mass of the ith asteroid respectively. If the asteroids were radially symmetric and uniformly dense, then by a volume argument, we could deduce that $y_i = \frac{4\pi}{3}\left(\frac{x_i}{2}\right)^3$. The asteroids however, are not radially symmetric nor uniformly dense, but we still suspect that x, y exhibit a cubic relationship, i.e. $y = p(x)$ where p is a cubic polynomial. Using the data $\left\{(x_i, y_i)\right\}_{i=1}^n$ in `prob3data.csv`, find the coefficients c_0, c_1, c_2, c_3 so that $p(x) := c_0 + c_1 x + c_2 x^2 + c_3 x^3$ is the least squares cubic estimator of y from x.

Solution: We can find c_0, c_1, c_2, c_3 by solving a least squares minimization problem $\min_c \|Ac - y\|_2^2$ where

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ 1 & x_3 & x_3^2 & x_3^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}, c = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}.$$

The optimal coefficients c can then be found using the normal equation $c = (A^\top A)^{-1} A^\top y$.

Problem 4: Gradient Methods

Recall the polynomial fitting approach from Problem 3. Suppose we want a solution that is robust to outliers. One way to do this is to replace the ℓ_2 norm in least squares with an ℓ_1 norm, where for a vector $x \in \mathbb{R}^n$, its ℓ_1 norm is given by $\|x\|_1 := \sum_{i=1}^n |x_i|$. This gives rise to the following optimization problem:

$$\min_x \|Ax - b\|_1 . \tag{1}$$

One common technique for optimization is called Gradient Descent which uses the function's derivative to iteratively reduce the objective value. Given a function f and an initial starting point x_0, Gradient descent produces a sequence of iterates $x_1, x_2, ...$ until convergence according to the following rule:

$$x_{k+1} = x_k - \alpha \nabla f(x)$$

where $\alpha \geq 0$ is the step size. Using the same A, b from Problem 3, implement a Gradient Descent algorithm to solve (1).

Solution: To implement gradient descent we need to compute the gradient of $\|Ax - b\|_1$. The jth entry of $\nabla_x \|Ax - b\|_1$ is $\frac{\partial \|Ax-b\|_1}{\partial x_j}$, so we can compute the gradient as follows:

$$\begin{aligned} \|Ax - b\|_1 &= \sum_{i=1}^n \left| \left(\sum_{j=1}^n A_{ij} x_j \right) - b_i \right| \\ \implies \frac{\partial \|Ax - b\|_1}{\partial x_j} &= \sum_{i=1}^n A_{ij} \text{sign} \left(\left(\sum_{j=1}^n A_{ij} x_j \right) - b_i \right) . \end{aligned}$$

Learning goals for this problem set:

Problem 1: To review unconstrained convex optimization.

Problem 2: To review stability analysis of discrete linear systems.

Problem 3: To review linear regression techniques and applications

Problem 4: To review first order optimization methods.