Online and Reinforcement Learning (2025)
Home Assignment 3

Davide Marchi 777881

Contents

1 Direct Policy Search

1.1 Multi-variate normal distribution

In this exercise we use the notation

$$N(m, C)$$

to denote the multivariate normal distribution with mean $m \in \mathbb{R}^n$ and covariance matrix $C \in \mathbb{R}^{n \times n}$. In particular, $N(0, I)$ denotes the standard normal distribution in $\mathbb{R}^n$. The notation here is consistent with that used in the Monte Carlo methods presentation (see `3.1 RL-MonteCarlo.pdf`).

1.

Let $a \in \mathbb{R}^n$ be a nonzero vector and consider the matrix

$$C = aa^T.$$

(a) Rank of $C = aa^T$

For any $x \in \mathbb{R}^n$ we have

$$Cx = aa^T x = a\,(a^T x).$$

Since $a^T x$ is a scalar, it follows that Cx is always a scalar multiple of a. In other words, the image (or column space) of C is contained in $\text{span}\{a\}$. Since $a \neq 0$, this is a one-dimensional subspace. Hence,

$$\text{rank}(C) = 1.$$

(b) Eigenvector and Eigenvalue of $C = aa^T$

We next show that a is an eigenvector of C. Indeed,

$$C\,a = aa^T a = a\,(a^T a) = \|a\|^2\, a.$$

Thus, a is an eigenvector corresponding to the eigenvalue

$$\lambda = \|a\|^2.$$

(c) Maximum Likelihood for a One-Dimensional Normal Distribution

Consider the family of one-dimensional normal distributions with zero mean and variance σ^2, that is,

$$N(0, \sigma^2).$$

The probability density function (pdf) is given by

$$p(a \mid \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{a^2}{2\sigma^2}\right).$$

For a single observation $a \in \mathbb{R}$, the likelihood function is

$$L(\sigma^2) = p(a \mid \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{a^2}{2\sigma^2}\right).$$

It is more convenient to maximize the logarithm of the likelihood:

$$\ell(\sigma^2) = \log L(\sigma^2) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{a^2}{2\sigma^2}.$$

Differentiate $\ell(\sigma^2)$ with respect to σ^2:

$$\frac{d\ell}{d\sigma^2} = -\frac{1}{2\sigma^2} + \frac{a^2}{2(\sigma^2)^2}.$$

Setting the derivative equal to zero, we obtain

$$-\frac{1}{2\sigma^2} + \frac{a^2}{2(\sigma^2)^2} = 0 \quad \Longrightarrow \quad \frac{a^2 - \sigma^2}{2(\sigma^2)^2} = 0.$$

Thus,

$$a^2 - \sigma^2 = 0 \quad \Longrightarrow \quad \sigma^2 = a^2.$$

This shows that the likelihood of generating $a \in \mathbb{R}$ is maximized when $\sigma^2 = a^2$.

2 Off-Policy Optimization in RiverSwim

3 Reward Shaping