**CMPUT 652: Reinforcement Learning with Robots** **Fall 2019**

# Lecture 3: September 10

*Instructor: Rupam Mahmood* *Scribe: Shivam Garg*

**Disclaimer**: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

## 3.1 Introduction

The use of machine learning to solve problems involves the following three broad steps:

- Approximating the ideal function (or the function to learn) with a parameterized function
- Defining the learning problem in terms of an objective function
- Deriving the learning methods from first principles

In the following sections, we will discuss each of these processes in more detail.

### 3.1.1 Formulating a parameterized function

A function $f$ is defined as a mapping from different inputs $X$ to their corresponding outputs $Y$. In our setting, we wish to learn an "ideal" function $f^*$ which is generating the data $X, Y$. Further, we wish to learn the ideal function using this data. One possible way to solve this problem is to fix a function $h = g(X)$ using human intuition.

Another approach, using Machine Learning, is to define a class or a space of functions $F_\theta = f(X;\theta)$ parameterized by some $\theta \in \Theta$. We then try to find the optimal parameter $\theta^*$ for which the difference between the ideal function $f^*$ and the parameterized function $F_{\theta^*} = f(X;\theta^*)$ is least. As illustrated in Fig. 3.1, it is possible to have a case where $F_\theta$ cannot represent the ideal function; in these cases the closest match is used instead.

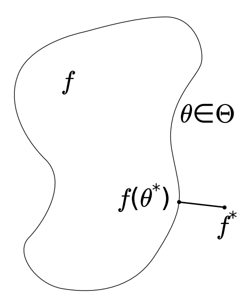


Figure 3.1: A learning scenario where the space of parameterized functions doesn't contain the ideal function $f^*$. In such cases, the function closest to $f^*$ is found instead.

### 3.1.2 Defining an objective function

After formulating a parameterized function, we need to define an objective in order to evaluate the parameterized function. The optimal parameter $\theta^*$ minimizes this objective function. There can be two types of objective functions:

1. Exact objective function: The complete form for an exact objective function is known, and there is no randomness associated with it. Such functions are mostly not available for real–world problems. However, they can be formulated for toy–problems and are useful for testing optimization algorithms. We represent exact objective functions as $\upsilon(\theta)$, where $\theta$ is the learnable parameter.

2. Sample–based objective function: These functions are noisy evaluations of the parameterized function based on observed data samples. We represent it as $\ell(X, F_\theta, Z)$, where $X$ is the input data, $F_\theta$ is the parameterized function, and $Z$ is the output of this parameterized function on input data. We expect the sample–based objective functions to be unbiased estimators of their corresponding exact functions: $\mathbb{E}[\ell(X, F_\theta, Z)] = \upsilon(\theta)$.

### 3.1.3 Learning Methods

Once the objective function has been formulated, the optimal parameter $\theta^*$ is found by minimizing the objective function: $\theta^* = \arg\min_{\theta\in\Theta} \upsilon(\theta)$, or $\theta^* = \arg\min_{\theta\in\Theta} \ell(X, F_\theta, Z)$. One possible way to solve this problem is by learning the parameters incrementally:

$$\theta_{t+1} = \theta_t + \Delta\theta_t, \tag{3.1}$$

where the parameter at each subsequent iteration decreases the value of the objective function. Two such incremental learning rules are Gradient descent: $\Delta\theta_t = -\alpha_t \nabla_\theta \upsilon(\theta)$ and Stochastic Gradient Descent: $\Delta\theta_t = -\alpha_t \nabla_\theta \ell(X, F_\theta, Z)$, where $\alpha_t$ is the step–size at timestep $t$ in both the cases.

#### Tabular Methods

Instead of using parameterized functions to approximate the ideal function, an alternative is to use tabular representation. In this case, the function outputs a fixed value for each entry of the table and in effect has no input (so the table can be viewed as being composed of multiple functions). We represent the tabular function as $f(\theta) = \theta$. The ideal function in this case would be $f^*$ (basically a constant) and we try to find $\theta^*$ such that $f(\theta^*)$ is closest to $f^*$.

## 3.2 Deriving squared loss from Maximum Likelihood Estimate

Consider a situation where we observe data sampled from a Normal distribution, with our goal being to estimate the mean of this Normal distribution using the observed data. Assume we observe the random variable $Y \sim \mathcal{N}(\theta^*, \sigma)$, with $\theta^*$ being the true mean of the underlying Normal distribution and $\sigma$ its standard deviation.

In using Maximum Likelihood Estimate, we try to maximize the log probability of drawing the observation given the mean $\theta$. That is

$$\begin{aligned}\theta^* &= \arg\max_{\theta\in\Theta}\ \log p(\{Y_i\}_{i=1}^N \mid \theta) &(3.2)\\ &= \arg\max_{\theta\in\Theta}\ \log \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y_i-\theta)^2}{2\sigma^2}}\\ &= \arg\max_{\theta\in\Theta}\ \left[-\frac{N}{2}\log(2\pi\sigma^2) - \sum_{i=1}^N \frac{(Y_i-\theta)^2}{2\sigma^2}\right]\\ &= \arg\min_{\theta\in\Theta}\ \sum_{i=1}^N (Y_i-\theta)^2, &(3.3)\end{aligned}$$

which is the sample–based batched loss function $\ell(F_\theta, \{Y_i\}_{i=1}^t) = \sum_{i=1}^N (Y_i-\theta)^2$.

We can use this loss function to calculate the exact form for the stochastic gradient descent update: $\Delta\theta_t = -\alpha\nabla_\theta \ell(F_\theta, Y_t) = 2\alpha_t(Y_t-\theta)$. The resulting update equation is

$$\theta_{t+1} = \theta_t + 2\alpha_t(Y_t - \theta_t). \tag{3.4}$$

### 3.2.1 Method of Averages

Our problem is to obtain the optimal parameter $\tilde{\theta}_{t+1}$ which minimizes the loss function $\ell(F_\theta, \{Y_i\}_{i=1}^t) = \sum_{i=1}^N (Y_i-\theta)^2$; i.e. $\tilde{\theta}_{t+1} = \arg\min_{\theta\in\Theta} \ell(F_\theta, \{Y_i\}_{i=1}^t)$. Instead of using the incremental stochastic gradient descent update, we can directly obtain the optimal parameter $\tilde{\theta}_{t+1}$ by using the complete batch of data. To do so, we set the derivative of the loss with respect to $\theta$ as zero and then solve for the optimal $\theta$.

$$\begin{aligned}&\frac{\partial}{\partial\tilde{\theta}_{t+1}} \sum_{i=1}^t (Y_i - \tilde{\theta}_{t+1})^2 = 0\\ \text{or}\quad &\sum_{i=1}^t 2(Y_i - \tilde{\theta}_{t+1})(-1) = 0\\ \text{or}\quad &\tilde{\theta}_{t+1} = \frac{1}{t}\sum_{i=1}^t Y_i. &(3.5)\end{aligned}$$

From this equation, we can see that the optimal parameter which minimizes the batched least squares error is the sample average of that batch. It is possible to compute this sample average in an incremental fashion too:

$$\tilde{\theta}_{t+1} = \tilde{\theta}_t + \frac{1}{t}(Y_t - \tilde{\theta}_t). \tag{3.6}$$

Yet another incremental form to compute the sample average is to maintain the sums for the numerator and denominator separately: $\tilde{\theta}_{t+1} = \frac{N_{t+1}}{D_{t+1}}$, with $N_{t+1} = N_t + Y_t$ and $D_{t+1} = D_t + 1$. All the three methods presented above for calculating the sample average would have different numerical stabilities. Further, note that Eq. 3.6 can be obtained from the stochastic gradient descent update (Eq. 3.4) by putting $\alpha_t = \frac{1}{2t}$.

#### 3.2.1.1 Constant step–size update

Another form of update uses a constant–step size $\alpha_t = \frac{\alpha}{2}$ in Eq. 3.4, which leads to an exponentially weighted sample average:

$$\begin{aligned}
\theta_{t+1} &= \theta_t + \alpha(Y_t - \theta_t) \\
&= (1-\alpha)\theta_t + \alpha Y_t \\
&\vdots \\
&= (1-\alpha)^{t+1}\theta_0 + \alpha(1-\alpha)^t Y_0 + \cdots + \alpha(1-\alpha)Y_{t-1} + \alpha Y_t \\
&= (1-\alpha)^{t+1}\theta_0 + \sum_{i=0}^{t} \alpha(1-\alpha)^{t-i} Y_i. \qquad (3.7)
\end{aligned}$$

## 3.3 Bias and Variance of different estimates

We first show the bias–variance decomposition of the exact objective function, in this case the Mean Squared Error (MSE), and then proceed to calculate the bias and variance of the sample average method and the exponential average method.

$$\begin{aligned}
v(\tilde{\theta}_{t+1}) &= \mathbb{E}[(Y - \tilde{\theta}_{t+1})^2] \qquad (3.8) \\
&= \mathbb{E}[(\theta^* - \tilde{\theta}_{t+1} + Y - \theta^*)^2] \qquad (3.9) \\
&= \mathbb{E}[(\theta^* - \tilde{\theta}_{t+1})^2 + (Y - \theta^*)^2 + 2(\theta^* - \tilde{\theta}_{t+1})(Y - \theta^*)] \qquad (3.10) \\
&= \mathbb{E}[(\theta^* - \tilde{\theta}_{t+1})^2] + \underbrace{\mathbb{E}[(Y - E[Y])^2]}_{\text{Variance of outcome}} + \underbrace{\mathbb{E}[2(\theta^* - \tilde{\theta}_{t+1})(Y - \theta^*)]}_{0} \qquad (3.11) \\
&= \mathbb{E}[(\theta^* - \mathbb{E}[\tilde{\theta}_{t+1}] + \mathbb{E}[\tilde{\theta}_{t+1}] - \tilde{\theta}_{t+1})^2] + \mathrm{Var}(Y) \\
&= \mathbb{E}[(\theta^* - \mathbb{E}[\tilde{\theta}_{t+1}])^2] + \underbrace{\mathbb{E}[(\mathbb{E}[\tilde{\theta}_{t+1}] - \tilde{\theta}_{t+1})^2]}_{\text{Variance of estimate}} + 2\underbrace{\mathbb{E}[(\theta^* - \mathbb{E}[\tilde{\theta}_{t+1}])(\mathbb{E}[\tilde{\theta}_{t+1}] - \tilde{\theta}_{t+1})]}_{0} + \mathrm{Var}(Y) \\
&= \underbrace{(\theta^* - \mathbb{E}[\tilde{\theta}_{t+1}])^2}_{\text{Bias}^2} + \mathrm{Var}(\tilde{\theta}_{t+1}) + \mathrm{Var}(Y). \qquad (3.12)
\end{aligned}$$

### 3.3.1 Bias and Variance of Sample Average Method

We now calculate the bias of the sample average estimate:

$$\begin{aligned}
\mathrm{Bias}(\tilde{\theta}_{t+1}) &= \mathbb{E}[\tilde{\theta}_{t+1}] - \theta^* \\
&= \mathbb{E}\left[\frac{1}{t}\sum_{i=1}^{t} Y_i\right] - \theta^* \\
&= \frac{1}{t} \cdot t\,\mathbb{E}[Y] - \theta^* \\
&= \theta^* - \theta^* \\
&= 0, \qquad (3.13)
\end{aligned}$$

where in the second line we replaced $\tilde{\theta}_{t+1}$ with the sample average from Eq. 3.5 and in the fourth line we replaced $\mathbb{E}[Y]$ with $\theta^*$ since $Y \sim \mathcal{N}(\theta^*, \sigma)$. From this equation, we see that the sample average is an unbiased estimator of the true sample mean.

The variance of the sample average estimate is:

$$\begin{aligned}
\mathrm{Var}(\tilde{\theta}_{t+1}) &= \mathbb{E}[\tilde{\theta}_{t+1}^2] - \mathbb{E}[\tilde{\theta}_{t+1}]^2 \\
&= \mathbb{E}\left[\left(\frac{1}{t}\sum_{i=1}^{t} Y_i\right)^2\right] - \theta^{*2} \\
&= \frac{1}{t^2}\,\mathbb{E}\left[\sum_{i=1}^{t} Y_i^2 + \sum_{\substack{i=1 \\ i\neq j}}^{t}\sum_{j=1}^{t} Y_i Y_j\right] - \theta^{*2} \\
&= \frac{1}{t^2}\sum_{i=1}^{t}\mathbb{E}[Y_i^2] + \frac{1}{t^2}\sum_{\substack{i=1 \\ i\neq j}}^{t}\sum_{j=1}^{t}\mathbb{E}[Y_i]\,\mathbb{E}[Y_j] - \theta^{*2} \\
&= \frac{1}{t^2}\cdot t(\theta^{*2} + \sigma^2) + \frac{t(t-1)}{t^2}\theta^{*2} - \theta^{*2} \\
&= \frac{\sigma^2}{t}, \qquad (3.14)
\end{aligned}$$

where in the second line we used the result obtained in the previous paragraph that $\mathbb{E}[\tilde{\theta}_{t+1}] = \theta^*$, in the fourth line we split the expectation operator over the product $Y_i Y_j$ because both of them are sampled i.i.d. from a normal distribution, and in the fifth line we replaced $\mathbb{E}[Y_i^2]$ with $(\theta^{*2} + \sigma^2)$ using the definition of variance. We observe that the variance of the sample average asymptotically goes to zero as the number of samples used to calculate the average increase.

### 3.3.2 Bias and Variance of Constant Step–size Estimate

The constant step–size estimate is $\theta_{t+1} = (1-\alpha)^{t+1}\theta_0 + \sum_{i=0}^{t}\alpha(1-\alpha)^{t-i}Y_i$ from Eq. 3.7. We now calculate the bias of this estimate:

$$\begin{aligned}
\mathrm{Bias}(\theta_{t+1}) &= \mathbb{E}[\theta_{t+1}] - \theta^* \\
&= \mathbb{E}\left[(1-\alpha)^{t+1}\theta_0 + \sum_{i=0}^{t}\alpha(1-\alpha)^{t-i}Y_i\right] - \theta^* \\
&= (1-\alpha)^{t+1}\theta_0 + \sum_{i=0}^{t}\alpha(1-\alpha)^{t-i}\,\mathbb{E}[Y_i] - \theta^* \\
&= (1-\alpha)^{t+1}\theta_0 + \alpha\theta^*\sum_{i=0}^{t}(1-\alpha)^{t-i} - \theta^* \\
&= (1-\alpha)^{t+1}\theta_0 + \alpha\theta^*\frac{1-(1-\alpha)^{t+1}}{1-(1-\alpha)} - \theta^* \\
&= (1-\alpha)^{t+1}(\theta_0 - \theta^*). \qquad (3.15)
\end{aligned}$$

If we assume that $-1 < 1-\alpha < 1$, that is, $0 < \alpha < 2$, then we see that the bias of the constant–step size estimate asymptotically goes to zero as the number of samples observed increases.

The variance calculation is:

$$
\begin{aligned}
\mathrm{Var}(\theta_{t+1}) &= \mathrm{Var}\left((1-\alpha)^{t+1}\theta_0 + \sum_{i=0}^{t} \alpha(1-\alpha)^{t-i} Y_i\right) \\
&= \mathrm{Var}((1-\alpha)^{t+1}\theta_0) + \sum_{i=0}^{t} (\alpha(1-\alpha)^{t-i})^2 \ \mathrm{Var}(Y_i)) \\
&= 0 + \sigma^2\alpha^2 \sum_{i=0}^{t} (1-\alpha)^{2(t-i)} \\
&= \sigma^2\alpha^2 \frac{1-(1-\alpha)^{2(t+1)}}{1-(1-\alpha)^2} \\
&= \alpha \frac{1-(1-\alpha)^{2(t+1)}}{2-\alpha}\sigma^2, \qquad (3.16)
\end{aligned}
$$

where because of the fact that $\theta_0$ and all the $Y_i$s are independent of each other, we use the rules $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$, $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$, and $\mathrm{Var}(a) = 0$ with $X$ and $Y$ being independent random variables and $a$ a constant. Again if we assume that $0 < \alpha < 2$, the $\mathrm{Var}(\theta_{t+1})$ asymptotically goes to $\frac{\alpha}{2-\alpha}\sigma^2$ as the value of $t$ increases.