AIL 722: Reinforcement Learning

Lecture 25: Experience Replay

Raunak Bhattacharyya

YARDI SCHOOL OF ARTIFICIAL INTELLIGENCE
INDIAN INSTITUTE OF TECHNOLOGY DELHI

Recap

- Online Q-Learning

- Exploration

Outline

- Serial correlation
- Experience replay
- Online Q-learning with replay buffer

Online Q-Learning

1. Take some action a_i and obtain (s_i, a_i, s'_i, r_i)

2. $y_i = r(s_i, a_i) + \gamma \cdot \max_{a'_i} Q(s'_i, a'_i)$

3. $\phi \longleftarrow \phi - \alpha \cdot \frac{dQ_\phi}{d\phi}(s_i, a_i) \cdot \left(Q_\phi(s_i, a_i) - y_i \right)$

Special case: Gradient step on tabular Q

A Problem

1. Take some action a_i and obtain (s_i, a_i, s'_i, r_i)

2. $\phi \longleftarrow \phi - \alpha \cdot \frac{dQ_\phi}{d\phi}(s_i, a_i) \cdot \left(Q_\phi(s_i, a_i) - [r(s_i, a_i) + \gamma \cdot \max_{a'_i} Q(s'_i, a'_i)] \right)$

Correlated samples

Set $\phi \longleftarrow \arg\min_\phi \sum_i \frac{1}{2} \left\| Q_\phi(s_i, a_i) - y_i \right\|^2$

Serial Correlation: Impact on Regression

$$\mathbb{E}(\epsilon_t) = 0$$

$$y_t = X_t\beta + \epsilon_t$$

$$\mathrm{Var}(\epsilon_t) = \sigma^2$$

$$\mathbb{E}(\epsilon_t\epsilon_s) = 0 \quad \text{for} \quad t \neq s$$

$$\hat{\beta}_{\mathrm{OLS}} = (X'X)^{-1}X'y$$

$$\hat{\beta}_{\mathrm{OLS}} = (X'X)^{-1}X'(X\beta + \epsilon)$$

$$\hat{\beta}_{\mathrm{OLS}} = \beta + (X'X)^{-1}X'\epsilon$$

Bias of the Estimator

$$\hat{\beta}_{\text{OLS}} = \beta + (X'X)^{-1}X'\epsilon$$

$$\text{Bias}(\hat{\beta}_{\text{OLS}}) = \mathbb{E}(\hat{\beta}_{\text{OLS}}) - \beta$$

$$\begin{aligned}
\mathbb{E}(\hat{\beta}_{\text{OLS}}) &= \mathbb{E}\left(\beta + (X'X)^{-1}X'\epsilon\right) \\
&= \beta + (X'X)^{-1}X'\mathbb{E}(\epsilon) \\
&= \beta
\end{aligned}$$

$$\text{Bias}(\hat{\beta}_{\text{OLS}}) = \mathbb{E}(\hat{\beta}_{\text{OLS}}) - \beta = 0$$

Variance of the Estimator

$$\hat{\beta}_{\mathrm{OLS}} = \beta + (X'X)^{-1}X'\epsilon$$

$$\mathrm{Var}(\hat{\beta}_{\mathrm{OLS}}) = \mathbb{E}\left[(\hat{\beta}_{\mathrm{OLS}} - \mathbb{E}(\hat{\beta}_{\mathrm{OLS}}))(\hat{\beta}_{\mathrm{OLS}} - \mathbb{E}(\hat{\beta}_{\mathrm{OLS}}))'\right]$$

Substitute $\hat{\beta}_{\mathrm{OLS}} = \beta + (X'X)^{-1}X'\epsilon$ and $\mathbb{E}(\hat{\beta}_{\mathrm{OLS}}) = \beta$:

$$\begin{aligned}\mathrm{Var}(\hat{\beta}_{\mathrm{OLS}}) &= \mathbb{E}\left[\left((X'X)^{-1}X'\epsilon\right)\left((X'X)^{-1}X'\epsilon\right)'\right] \\ &= (X'X)^{-1}X'\mathbb{E}(\epsilon\epsilon')X(X'X)^{-1}\end{aligned}$$

Since $\mathbb{E}(\epsilon\epsilon') = \sigma^2 I_n$, we get:

$$\mathrm{Var}(\hat{\beta}_{\mathrm{OLS}}) = \sigma^2(X'X)^{-1}$$

With Serial Correlation

$$y_t = X_t\beta + \epsilon_t \qquad \mathbb{E}(\epsilon_t) = 0$$

First-order autoregressive structure

$$\epsilon_t = \rho\epsilon_{t-1} + u_t$$

where:

- ρ is the **autocorrelation coefficient** (with $|\rho| < 1$),
- u_t is a white noise error term with $\mathbb{E}(u_t) = 0$ and $\mathrm{Var}(u_t) = \sigma^2$,
- t ranges from 1 to n, where n is the **total number of observations**

With Serial Correlation

$$y_t = X_t\beta + \epsilon_t \qquad \epsilon_t = \rho\epsilon_{t-1} + u_t$$

$$\Omega = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}$$

Diagonal terms are variance of each error term

Off-diag terms represent covariance between different time steps

With Serial Correlation

$$\epsilon_t = \rho\epsilon_{t-1} + u_t$$

$$\mathbb{E}(\hat{\beta}_{\text{OLS}}) = \beta$$

$$\text{Bias}(\hat{\beta}_{\text{OLS}}) = 0$$

$$\text{Var}(\hat{\beta}_{\text{OLS}}) = (X'X)^{-1}X'\Omega X(X'X)^{-1}$$

Variance of the OLS estimator is larger in the presence of serial correlation