FPT Education
FPT UNIVERSITY

# Linear Temporal Difference

# Objectives

- Understand that tabular TD is a special case of linear semi-gradient TD
- Understand the fixed point of linear TD learning

# Temporal Difference Update

- TD update with linear function approximation
  - Semi gradient TD adjusts the weights and the direction of the TD air times the gradient of the approximate value function with respect to the weights.

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha\delta_t \nabla \hat{v}(S_t, \mathbf{w})$$

$$\delta_t \doteq R_{t+1} + \gamma\hat{v}(S_{t+1}, \mathbf{w}) - \hat{v}(S_t, \mathbf{w})$$

  - In the linear case, the gradient of the approximate value for a state is just the feature vector for that state

$$\hat{v}(S_t, \mathbf{w}) \doteq \mathbf{w}^T\mathbf{x}(S_t)$$

$$\nabla \hat{v}(S_t, \mathbf{w}) = \mathbf{x}(S_t)$$

# Temporal Difference Update

- The update for semi gradient TD would look like:

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha \delta_t \nabla \hat{v}(S_t, \mathbf{w}) \qquad \delta_t \doteq R_{t+1} + \gamma \hat{v}(S_{t+1}, \mathbf{w}) - \hat{v}(S_t, \mathbf{w})$$

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha \delta_t \mathbf{x}(S_t)$$

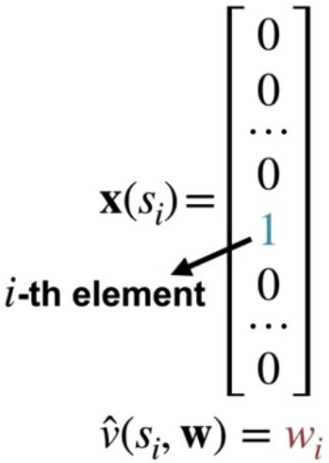
$$\hat{v}(S_t, \mathbf{w}) \doteq \mathbf{w}^T \mathbf{x}(S_t)$$

$$\nabla \hat{v}(S_t, \mathbf{w}) = \mathbf{x}(S_t)$$

  - The weight is updated in the direction of the feature vector times the TD air.
  - If a feature is large, then the corresponding weight can have a large impact on the prediction.
  - If the feature is zero, then that weight has no impact on the prediction and the gradient is therefore zero.

# Temporal Difference Update

- Tabular TD is a special case of liner TD
  - We have a tabular state representation( Only one feature will be equal to one corresponding to the current state)




  - The value approximation for a state is equal to the weight associated with the current state
  - $$\mathbf{w} \leftarrow \mathbf{w} + \alpha[R_{t+1} + \gamma \hat{v}(S_{t+1}, \mathbf{w}) - \hat{v}(S_t, \mathbf{w})]\mathbf{x}(S_t)$$

tabular features

# Temporal Difference Update

- Tabular TD is a special case of liner TD
  - In the update, the feature vector X of ST selects a single weight associated with the current state.

$$w_i \leftarrow w_i + \alpha[R_{t+1} + \gamma \hat{v}(S_{t+1}, \mathbf{w}) - \hat{v}(S_t, \mathbf{w})]1$$

  - This weight is just the value estimate for the state. So this update corresponds to the tabular TD update we saw in a previous course.
  - We can use the same analysis to show that TD with state aggregation is also a special case of linear TD

# The Expected TD Update

- The TD update with linear function approximation
  - The value of a state is the inner product of the state features and the learn weight vector.
  - X_ t to mean the features associated with the state S_t.

$$\mathbf{w}_{t+1} \doteq \mathbf{w}_t + \alpha[R_{t+1} + \gamma\hat{v}(S_{t+1}, \mathbf{w}_t) - \hat{v}(S_t, \mathbf{w}_t)]\mathbf{x}_t \qquad \hat{v}(s, \mathbf{w}) \doteq \mathbf{w}^T\mathbf{x}(s)$$

# The Expected TD Update

- The TD update with linear function approximation
  - We can rewrite this using a bit of linear algebra.
  - We'll pull X_t into the brackets.
  - We then use the fact that taking the transpose of a scalar, leaves it unchanged.

$$\mathbf{w}_{t+1} \doteq \mathbf{w}_t + \alpha[R_{t+1} + \gamma\hat{v}(S_{t+1}, \mathbf{w}_t) - \hat{v}(S_t, \mathbf{w}_t)]\mathbf{x}_t \qquad \hat{v}(s, \mathbf{w}) \doteq \mathbf{w}^T\mathbf{x}(s)$$

$$= \mathbf{w}_t + \alpha[R_{t+1} + \gamma\mathbf{w}_t^T\mathbf{x}_{t+1} - \mathbf{w}_t^T\mathbf{x}_t]\mathbf{x}_t$$

$$= \mathbf{w}_t + \alpha[R_{t+1}\mathbf{x}_t - \mathbf{x}_t(\mathbf{x}_t - \gamma\mathbf{x}_{t+1})^T\mathbf{w}_t]$$

# The Expected TD Update

- The TD update with linear function approximation
  - The TD update can be rewritten as the expected update plus a noise term, and so is largely dominated by the behavior of the expected update.
  - The expected update characterizes the expected change in the weight from one time step to the next.
  - The expected TD update can be written as a vector b minus a matrix A times the current weight.

$$\mathbf{w}_{t+1} \doteq \mathbf{w}_t + \alpha[R_{t+1} + \gamma\hat{v}(S_{t+1}, \mathbf{w}_t) - \hat{v}(S_t, \mathbf{w}_t)]\mathbf{x}_t \qquad \hat{v}(s, \mathbf{w}) \doteq \mathbf{w}^T\mathbf{x}(s)$$

$$= \mathbf{w}_t + \alpha[R_{t+1} + \gamma\mathbf{w}_t^T\mathbf{x}_{t+1} - \mathbf{w}_t^T\mathbf{x}_t]\mathbf{x}_t$$

$$= \mathbf{w}_t + \alpha[R_{t+1}\mathbf{x}_t - \mathbf{x}_t(\mathbf{x}_t - \gamma\mathbf{x}_{t+1})^T\mathbf{w}_t]$$

$$\mathbb{E}[\Delta\mathbf{w}_t] = \alpha(\mathbf{b} - \mathbf{A}\mathbf{w}_t)$$

# The Expected TD Update

- The TD update with linear function approximation
  - The matrix A is defined in terms of an expectation over the features, while the vector b is defined in terms of the features and the reward.

$$\mathbf{w}_{t+1} \doteq \mathbf{w}_t + \alpha[R_{t+1} + \gamma \hat{v}(S_{t+1}, \mathbf{w}_t) - \hat{v}(S_t, \mathbf{w}_t)]\mathbf{x}_t \qquad \hat{v}(s, \mathbf{w}) \doteq \mathbf{w}^T \mathbf{x}(s)$$

$$= \mathbf{w}_t + \alpha[R_{t+1} + \gamma \mathbf{w}_t^T \mathbf{x}_{t+1} - \mathbf{w}_t^T \mathbf{x}_t]\mathbf{x}_t$$

$$= \mathbf{w}_t + \alpha[R_{t+1}\mathbf{x}_t - \mathbf{x}_t(\mathbf{x}_t - \gamma \mathbf{x}_{t+1})^T \mathbf{w}_t]$$

$$\mathbb{E}[\Delta \mathbf{w}_t] = \alpha(\mathbf{b} - \mathbf{A}\mathbf{w}_t)$$

$$\mathbf{b} \doteq \mathbb{E}[R_{t+1}\mathbf{x}_t] \qquad \mathbf{A} \doteq \mathbb{E}[\mathbf{x}_t(\mathbf{x}_t - \gamma \mathbf{x}_{t+1})^T]$$

# The Expected TD Update

- The TD Fixed point:
  - The weights are said to converge, when this expected TD update is zero.
  - We call this point W_TD.
  - If A is invertible, we can express this as W_TD equals A inverse b.
  - More generally, W_TD is a solution to this linear system.

$$\mathbb{E}[\Delta \mathbf{w}_{TD}] = \alpha(\mathbf{b} - \mathbf{A}\mathbf{w}_{TD}) = 0$$

$$\Longrightarrow \quad \mathbf{w}_{TD} = \mathbf{A}^{-1}\mathbf{b}$$

# The Expected TD Update

- The TD Fixed point:
  - It can be shown that TD minimizes an objective that is based on this A and b.
  - This objective extends the connection between TD and Bellman equations, to the function approximation setting.

$$\mathbb{E}[\Delta \mathbf{w}_{TD}] = \alpha(\mathbf{b} - \mathbf{A}\mathbf{w}_{TD}) = 0$$

$$\Longrightarrow \quad \mathbf{w}_{TD} = \mathbf{A}^{-1}\mathbf{b}$$

$$\mathbf{w}_{TD} \text{ minimizes } (\mathbf{b} - \mathbf{A}\mathbf{w})^T(\mathbf{b} - \mathbf{A}\mathbf{w})$$

# TD Fixed Poin

- The relating the TD Fixed Point and the Minimum of the Value Error

$$\overline{VE}(\mathbf{w}_{TD}) \leq \frac{1}{1-\gamma} \min_{\mathbf{w}} \overline{VE}(\mathbf{w})$$

  - The difference between the TD fixed point and the minimum value error solution can be large if Gamma is close to one.
  - If Gamma is very close to zero on the other hand, the TD fixed point is very close to the minimum value error solution.

# Summary

- Understand that tabular TD is a special case of linear semi-gradient TD
- Understand the fixed point of linear TD learning

FPT Education
FPT UNIVERSITY

# Q & A