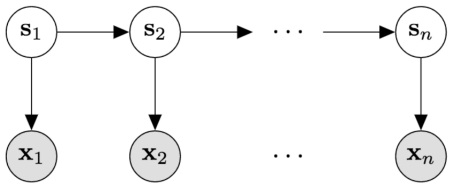
CS 181 Spring 2021 Section 11

HMMs, Kalman Filters, and MDP's

1 Hidden Markov Models

A Hidden Markov Model (HMM) is useful for inferring a sequence of unknown or hidden states from a corresponding sequence of observed evidence.

1.1 Graphical Model



Consider a sequence of one-hot encoded states $\mathbf{s}_1$,...,$\mathbf{s}_n$ where $\mathbf{s}_t \in \{S_k\}_{k=1}^{c}$, and a corresponding sequence of observations ($\mathbf{x}_1$,...,$\mathbf{x}_n$) where $\mathbf{x}_t \in \{O_j\}_{j=1}^{m}$. Each state can be one of c possible states, and each observation can be one of m possible observations. Note that N is the number of data points (each of which is a sequence), where n is the length of a sequence (assume all sequences are the same length).

1.2 Model Assumptions

HMMs are characterized by and allow us to reason about the following joint distribution

$$p(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \mathbf{x}_1, \ldots, \mathbf{x}_n) = p(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n) p(\mathbf{x}_1, \ldots, \mathbf{x}_n \mid \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)$$

However, it's not immediately obvious how we should optimize this model, and the following assumptions make this easier:

- The future hidden state is independent of past hidden states given the present (Markov Property):

  $p(\boldsymbol{s}_{t+1} \mid \boldsymbol{s}_1, \ldots \boldsymbol{s}_t, \mathbf{x}_1, \ldots, \mathbf{x}_t) = p(\boldsymbol{s}_{t+1} \mid \boldsymbol{s}_t)$

- Observations only depend on the present hidden state:

  $p(\mathbf{x}_t \mid \boldsymbol{s}_1, \ldots, \boldsymbol{s}_t, \mathbf{x}_1, \ldots, \mathbf{x}_{t-1}) = p(\mathbf{x}_t \mid \boldsymbol{s}_t)$

Notice that the above assumptions allow us to factor the joint as follows:

$$
\begin{aligned}
p(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \mathbf{x}_1, \ldots, \mathbf{x}_n) = \\
p(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n) p(\mathbf{x}_1, \ldots, \mathbf{x}_n \mid \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n) = \\
p(\boldsymbol{s}_1) \prod_{t=1}^{n-1} p(\boldsymbol{s}_{t+1} \mid \boldsymbol{s}_t) \prod_{t=1}^{n} p(\mathbf{x}_t \mid \boldsymbol{s}_t)
\end{aligned}
$$

1.3 Exercise: When to Use HMMs (Source: CMU)

For each of the following scenarios, is it appropriate to use a Hidden Markov Model? Why or why not? What would the observed data be in each case, and what would the hidden states capture?

1. Stock market price data
2. Recommendations on a database of movie reviews
3. Daily precipitation data in Boston
4. Optical character recognition for identifying words

1.4 Parameterization

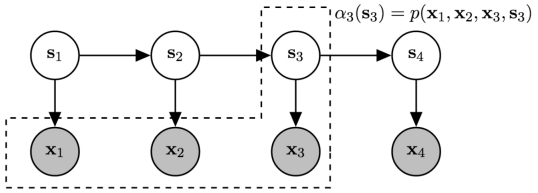
- $\boldsymbol{\theta} \in \mathbb{R}^c$: defines the prior distribution over initial hidden states
- $\mathbf{T} \in \mathbb{R}^{c \times c}$: transition matrix where T_{kj} is the probability of transitioning from S_k to S_j
- $\{\boldsymbol{\pi}\}_{k=1}^c$: conditional probabilities of observations given hidden states such that $p(\mathbf{x}_t = O_j | \mathbf{s}_t = S_k; \{\boldsymbol{\pi}\}) = \pi_{kj}$. $\forall k$ $\boldsymbol{\pi}_k \in \mathbb{R}^m$.

First, we need to estimate the parameters from the data, which we can do with a variant of EM. Then, with our trained HMM, we are able to perform several inference tasks on our data.

1.5 Forward-Backward Algorithm

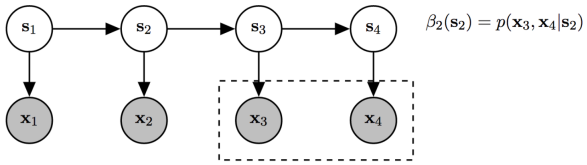
The HMM model is characterized by the joint distribution $p(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \mathbf{x}_1, \ldots, \mathbf{x}_n)$, which means that many of our training and inference tasks require marginalization to obtain conditionals. Thus, naive algorithms can be expensive (they require lots of nested summations over states), and we use EM instead. We define the recurrence relations $\alpha_t(\boldsymbol{s}_t)$ and $\beta_t(\boldsymbol{s}_t)$ in the E-Step:

- $\alpha_t(\mathbf{s}_t)$ represents the joint probability of observations $1, \ldots, t$ and state t. α_t can be defined in terms of α_{t-1}. We move **forwards** through the sequence to calculate the α's
- $\beta_t(\mathbf{s}_t)$ represents the joint probability of observations $t+1, \ldots, n$ conditioned on state t. β_t can be defined in terms of β_{t+1}. We move **backwards** through the sequence to calculate the β's.



$$\forall \mathbf{s}_3 : \quad \alpha_3(\mathbf{s}_3) = p(\mathbf{x}_3 \mid \mathbf{s}_3) \sum_{\mathbf{s}_2} p(\mathbf{s}_3 \mid \mathbf{s}_2) \alpha_2(\mathbf{s}_2)$$

(a) alpha



$$\forall \mathbf{s}_2 : \quad \beta_2(\mathbf{s}_2) = \sum_{s_3} p(\mathbf{s}_3 \mid \mathbf{s}_2) p(\mathbf{x}_3 \mid \mathbf{s}_3) \beta_3(\mathbf{s}_3)$$

(b) beta

Note that the probabilities we use for calculating α and β are given by the parameters that we fix in the E-Step.

$$\forall \mathbf{s}_t : \quad \alpha_t(\mathbf{s}_t) = \begin{cases} p(\mathbf{x}_t \mid \mathbf{s}_t) \sum_{\mathbf{s}_{t-1}} p(\mathbf{s}_t \mid \mathbf{s}_{t-1}) \alpha_{t-1}(\mathbf{s}_{t-1}) & \text{if } 1 < t \leq n \\ p(\mathbf{x}_1 \mid \mathbf{s}_1) p(\mathbf{s}_1) & \text{o.w.} \end{cases}$$

$$\forall \mathbf{s}_t : \quad \beta_t(\mathbf{s}_t) = \begin{cases} \sum_{\mathbf{s}_{t+1}} p(\mathbf{s}_{t+1} \mid \mathbf{s}_t) p(\mathbf{x}_{t+1} \mid \mathbf{s}_{t+1}) \beta_{t+1}(\mathbf{s}_{t+1}) & \text{if } 1 \leq t < n \\ 1 & \text{o.w.} \end{cases}$$

1.6 EM for HMMs

Given data points $\{\mathbf{x}^i\}_{i=1}^N$ defined by sequences $(x_1^i, \ldots, x_n^i)$ of length n represented as row vectors, we want to infer the parameters $\{\mathbf{T}, \boldsymbol{\theta}, \{\boldsymbol{\pi}_k\}\}$. Had we been given the true states, we could easily compute joint probability $p(\mathbf{x}^i, \mathbf{s}^i)$ and write the complete-data log likelihood, and maximize with respect to the parameters. Instead, we need to estimate state distributions and parameters iteratively.

1.6.1 Inference Patterns with $\boldsymbol{\alpha}, \boldsymbol{\beta}$

The following patterns are useful for inference with a trained HMM as well as during the E-Step:

- $\alpha_t(\mathbf{s}_t)\beta_t(\mathbf{s}_t) = p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{s}_t) \propto p(\mathbf{s}_t|\mathbf{x}_1, \ldots, \mathbf{x}_n)$
- joint of observations: $p(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \sum_{\mathbf{s}_t} \alpha_t(\mathbf{s}_t)\beta_t(\mathbf{s}_t)$ (for any t)
- smoothing: $p(\mathbf{s}_t \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n) \propto p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{s}_t) = \alpha_t(\mathbf{s}_t)\beta_t(\mathbf{s}_t)$
- prediction: $p(\mathbf{x}_{n+1} \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n) \propto \sum_{\mathbf{s}_n, \mathbf{s}_{n+1}} \alpha_n(\mathbf{s}_n) p(\mathbf{s}_{n+1} \,|\, \mathbf{s}_n) p(\mathbf{x}_{n+1} \,|\, \mathbf{s}_{n+1})$
- transition: $p(\mathbf{s}_t, \mathbf{s}_{t+1} \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n) \propto \alpha_t(\mathbf{s}_t) p(\mathbf{s}_{t+1} \,|\, \mathbf{s}_t) p(\mathbf{x}_{t+1} \,|\, \mathbf{s}_{t+1}) \beta_{t+1}(\mathbf{s}_{t+1})$

1.6.2 E-Step

The goal of the expectation step is to compute the expected values of the hidden states given a fixed set of parameters $\mathbf{w} = \{\mathbf{T}, \boldsymbol{\theta}, \{\boldsymbol{\pi}_k\}\}$. That is, we estimate the state distribution for $\mathbf{s}_1^i, \ldots, \mathbf{s}_n^i$ given $\mathbf{x}^i$.

Let the $c \times 1$ vector $\mathbf{q}_t^i = (q_{t1}^i, \ldots, q_{tc}^i)$ represent $\mathbf{x}^i$'s distribution over states for time t under the current parameters. Let $\mathbf{Q}_{t,t+1}^i$ be the $c \times c$ matrix of transition probabilities under the current parameters. Then

- α's and β's are defined in terms of fixed parameters.
- $\mathbf{q}$'s are defined in terms of α's and β's
- Calculate $q_{tk}^i = p(\mathbf{s}_t^i = S_k|\mathbf{x}^i; \mathbf{w})$ for all t and k (use smoothing eq. just above)
- Calculate $q_{t,t+1,k,\ell}^i = p(\mathbf{s}_t^i = S_k, \mathbf{s}_{t+1}^i = S_\ell|\mathbf{x}^i; \mathbf{w})$ (use transition eq. just above)

1.6.3 M-Step

Now we need to update our parameters to maximize the expected complete-data log likelihood $\mathbb{E}_{\mathbf{S}}[\ln p(\mathbf{x}, \mathbf{S}; \mathbf{w})]$. Applying the appropriate Lagrange multipliers and maximizing with respect to each of the parameters of interest, we recover the following update equations:

$$\hat{N}_{1k} = \sum_{i=1}^{N} q_{1k}^i \text{ (first period)} \quad \text{and more generally} \quad \hat{N}_k = \sum_{i=1}^{N} \sum_{t=1}^{n} q_{tk}^i \text{ (all periods)}$$

$$\hat{N}_{-nk} = \sum_{i=1}^{N} \sum_{t=1}^{n-1} q_{tk}^i \text{ (without last period)}$$

$$\hat{N}_{k\ell} = \sum_{i=1}^{N} \sum_{t=1}^{n-1} q^i_{t,t+1,k,\ell} \text{ (transitions)}$$

$$\hat{N}_{kj} = \sum_{i=1}^{N} \sum_{t=1}^{n} q^i_{tk} x^i_{tj} \text{ (observations)}$$

$$\hat{\theta}_k = \frac{\hat{N}_{1k}}{N} \quad \hat{\pi}_{kj} = \frac{\hat{N}_{kj}}{\hat{N}_k} \quad \hat{t}_{k\ell} = \frac{\hat{N}_{k\ell}}{\hat{N}_{-nk}}$$

1.7 Exercise: Parameter Estimation in Supervised HMMs

You are trying to predict the weather using an HMM. The hidden states are the weather of the day, which may be sunny or rainy, and the observable states are the color of the clouds, which can be white or gray. You have data on the weather and clouds from one sequence of four days (note: the hidden states are observed here):

Day	Weather	Clouds
1	Sunny	White
2	Rainy	Gray
3	Rainy	Gray
4	Sunny	Gray

1. Draw a graphical model representing the HMM.
2. Give the values of N, n, c and of the one-hot vectors $\mathbf{s}_1^1, \ldots, \mathbf{s}_4^1, \mathbf{x}_1^1, \ldots, \mathbf{x}_4^1$.
3. Estimate and interpret the values of the parameters $\boldsymbol{\theta}, \mathbf{T}, \{\boldsymbol{\pi}_k\}_{k=1}^c$ using the MLE estimators for the supervised HMM:

$$\hat{\theta}_k = \frac{N_{1k}}{N}, \quad \hat{t}_{kl} = \frac{N_{kl}}{N_{-nk}}, \quad \hat{\pi}_{kj} = \frac{N_{kj}}{N_k}$$

$$N_k = \sum_{i=1}^{N} \sum_{t=1}^{n} s_{tk}^i, \quad N_{1k} = \sum_{i=1}^{N} s_{1,k}^i, \quad N_{-nk} = \sum_{i=1}^{N} \sum_{t=1}^{n-1} s_{tk}^i$$

$$N_{kl} = \sum_{i=1}^{N} \sum_{t=1}^{n-1} s_{t,k}^i s_{t+1,l}^i, \quad N_{kj} = \sum_{i=1}^{N} \sum_{t=1}^{n} s_{tk}^i x_{tj}^i$$

1.8 Exercise: EM for HMMs

You are trying to model a toy's state using an HMM. At each time step, the toy can be active (state 1) or inactive (state 2), but you can only observe the color of the indicator light, which can be red (observation state 1) or green (observation state 2). You have collected data from one sequence:

Time	Light
1	Green
2	Red
3	Green

You initialize your EM with $\boldsymbol{\theta} = [\frac{1}{2}\ \frac{1}{2}]^\top, \mathbf{T} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}, \boldsymbol{\pi}_1 = [\frac{1}{4}\ \frac{3}{4}]^\top, \boldsymbol{\pi}_2 = [\frac{3}{4}\ \frac{1}{4}]^\top$.

1. Compute $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3$ for the forward-backward algorithm using the initial parameter values.
2. Refer to the definition of $\mathbf{q}_t^1$ in Section 1.6.2. Now, compute the values of $\mathbf{q}_1^1, \mathbf{q}_2^1$ using the α and β values.
3. Refer to the definition of $\mathbf{Q}_{t,t+1}^1$ in Section 1.6.2. Compute the value of $\mathbf{Q}_{1,2}^1$ using the α and β values.

During EM, at one point you obtain the following values after the E step:

$$\mathbf{q}_1^1 = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}^\top, \quad \mathbf{q}_2^1 = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix}^\top, \quad \mathbf{q}_3^1 = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}^\top$$

$$\mathbf{Q}_{1,2}^1 = \begin{bmatrix} \frac{1}{6} & \frac{1}{2} \\ \frac{1}{6} & \frac{1}{6} \end{bmatrix}, \quad \mathbf{Q}_{2,3}^1 = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} \\ \frac{1}{2} & \frac{1}{6} \end{bmatrix}$$

1. Use the above values to compute $\hat{N}_k, \hat{N}_{kl}, \hat{N}_{kj}$.
2. Complete the M step by updating the parameters $\boldsymbol{\theta}, \mathbf{T}, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2$.

2 Markov Decision Processes

A Markov Decision Process (MDP) is a framework for modeling an agent's actions in the world. It consists of:

1. A set of states S
2. A set of actions A
3. A reward function $r : S \times A \to \mathbb{R}$
4. A transition model $p(s'|s, a)$, $\forall s, s' \in S, a \in A$.

A *policy* π is a mapping from states to actions, i.e. $\pi : S \to A$.

2.1 Finite time horizon MDP

In the finite horizon setting, a policy may vary with the number of time periods remaining. $\pi_{(t)}$ denotes the policy with t time steps to go. T is the decision horizon. The value of a policy with t time steps to go is defined inductively to be:

$$V^{\pi}_{(t)}(s) = \begin{cases} r(s, \pi_{(1)}(s)) & \text{if } t = 1 \\ r(s, \pi_{(t)}(s)) + \sum_{s' \in S} p(s'|s, \pi_{(t)}(s)) V^{\pi}_{(t-1)}(s') & \text{o.w.} \end{cases} \tag{1}$$

The process of computing these values inductively, working from the end of the horizon to the present, is called *value iteration*. If we instead look forward in time, we are computing the expected value of the policy

$$V^{\pi}_{T}(s) = \mathbb{E}_{s_1, \ldots, s_T} \left[\sum_{t=0}^{T} r(s_t, \pi_{(T-t)}(s_t)) \right] \tag{2}$$

by induction, where $s_1 := s$. $V^{\pi}(s)$ is the MDP value function.

In an MDP, the general goal is to find an optimal policy by maximizing the expected reward under the policy, i.e. maximizing the value function. This is the planning problem.

2.2 Infinite Horizon MDP

Policy Evaluation We can also send $T \to \infty$, i.e. have an infinite time horizon. In that case, we need a discount factor $0 < \gamma < 1$, and we want to compute the value function

$$V^\pi(s) = \mathbb{E}_{s_1, s_2, \ldots}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, \pi(s_t))\right] \tag{3}$$

where $s_1 := s$, and the γ factor ensures convergence (assuming bounded rewards). In this setting, we only worry about stationary policies that don't vary with time. This is the policy evaluation problem; for any given policy π, we can find $V^\pi(s)$ by solving the system of linear equations

$$V^\pi(s) = r(s, \pi(s)) + \gamma \sum_{s' \in S} p(s'|s, \pi(s)) V^\pi(s') \tag{4}$$

These capture consistency about the value function. To solve this system, we can use Gaussian elimination, or simply iterate until convergence as in the finite horizon case.

Given a policy π and θ (small positive number), we find V^π iteratively as follows:

- Initialize: $V(s) = 0$ for all states s.
- Repeat
  - Update step:

$$V'(s) = r(s, \pi(s)) + \gamma \sum_{s' \in S} p(s'|s, \pi(s)) V(s'), \ \forall s \tag{5}$$

  - $\Delta = \max(|V' - V|)$
  - $V \leftarrow V'$

  until $\Delta < \theta$

Value Iteration Suppose we have an optimal policy π^*. This satisfies the following set of equations known as the Bellman equations:

$$V^*(s) = \max_{a \in A}\left[r(s, a) + \gamma \sum_{s' \in S} p(s'|s, a) V^*(s')\right] \tag{6}$$

where $V^* \triangleq V^{\pi^*}$. Assuming we know V^*, we can read off the optimal policy by setting

$$\pi^*(s) = \arg\max_{a \in A} \left[r(s,a) + \gamma \sum_{s' \in S} p(s'|s,a) V^*(s') \right] \tag{7}$$

In order to find V^*, we can use value iteration:

- Initialize: $V(s) = 0$ for all states s.
- Update step (Bellman operator):

$$V'(s) = \max_{a \in A} \left[r(s,a) + \gamma \sum_{s' \in S} p(s'|s,a) V(s') \right], \ \forall s \tag{8}$$

- $V \leftarrow V'$

where we iterate until convergence of V, which is guaranteed. With our converged V, we can then find π^* as in Equation 7.

Policy Iteration Another approach to planning is called *policy iteration*. To do policy iteration, we evaluate a proposed policy π by finding V^π as in Equation 4. This is Evaluation step (E step). Then, we do a policy improvement step (I step) by the equation

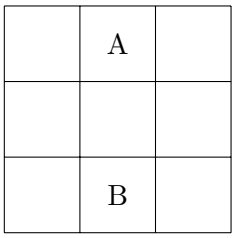
$$\pi'(s) \leftarrow \arg\max_{a \in A} \left[r(s,a) + \gamma \sum_{s' \in S} p(s'|s,a) V^\pi(s') \right], \quad \forall s \tag{9}$$

We repeat the E and I steps until the policy π converges (stops changing).

Note that policy iteration takes more computation per iteration, but tends to converge faster in practice.

2.3 Exercise: Markov Decision Process

(Sutton & Barto 2012) Consider an MDP on the following grid:



At each square, we can go left, right, up, or down. Normally we get a reward of 0 from moving, but if we attempt to move off the grid, we get a reward of -1 and stay where we are. Also, if we move onto square A, we get a reward of 10 and are teleported to square B.

Suppose our actions also fail with probability 0.5, i.e. with probability 0.5 we stay on the current square. Also suppose our MDP is infinite horizon, and take $\gamma = 0.9$ to be the discount factor.

1. **Defining the MDP** Identify the states S, actions A, rewards, and transition probabilities $p(s'|s, a)$ in this problem.
2. **Policy Evaluation** Suppose π is the policy where we always choose to go right. Write the equations to find the values $V^\pi(s)$.
3. **Value Iteration** Write the second iteration of value iteration, i.e. starting by initializing $V(s) = \max_{a \in A} r(s, a)$.
4. **Policy Iteration** Write the first iteration of policy iteration, starting with $V^\pi(s) = 0$ for all s. (We could also initialize a policy, and do the Evaluation step to get started.)

3 Kalman Filters (Bonus Material)

Now consider the following dynamical system model:

$$z_{t+1} = \Phi z_t + \epsilon_t$$

$$x_t = A z_t + \gamma_t$$

where z are the hidden variables and x are the observed measurements. Φ and A are known constants, while ϵ and γ are random variables drawn from the following normal distributions:

$$\epsilon_t \sim \mathcal{N}(\mu_\epsilon, \sigma_\epsilon^2)$$

$$\gamma_t \sim \mathcal{N}(\mu_\gamma, \sigma_\gamma^2)$$

This is called a (one-dimensional) linear Gaussian state-space model. It is closely related to an HMM – try drawing out the graphical model! – but here the hidden states and the observations are now continuous and normally distributed. Linear Gaussian state-space models have convenient mathematical properties and can be used to describe noisy measurements of a moving object (e.g. missiles, rodents, hands), market fluctuations, etc.

The Kalman filter is an algorithm to perform filtering in linear Gaussian state-space models, i.e. to find the distribution of z_t given observations $x_1, ..., x_t$. The distribution of $z_t \,|\, x_1, ..., x_s$ will be $\mathcal{N}(\mu_{t|s}, \sigma_{t|s}^2)$. If we start with $\mu_{t-1|t-1}$ and $\sigma_{t-1|t-1}^2$, the algorithm tells us to

1. Define the distribution of $z_t \,|\, x_1, ..., x_{t-1}$ by computing $\mu_{t|t-1}$ and $\sigma_{t|t-1}^2$. This is called the prediction step.

2. Define the distribution of $z_t \,|\, x_1, ..., x_t$ by computing $\mu_{t|t}$ and $\sigma_{t|t}^2$. This is called the update step.

The Kalman filter alternates between prediction and update steps, assimilating observations one at a time. It requires one forward pass through the data, and is analogous to obtaining the α's in an HMM.