Statistical Machine Learning

Lecture 06 Extra: Expectation Maximization

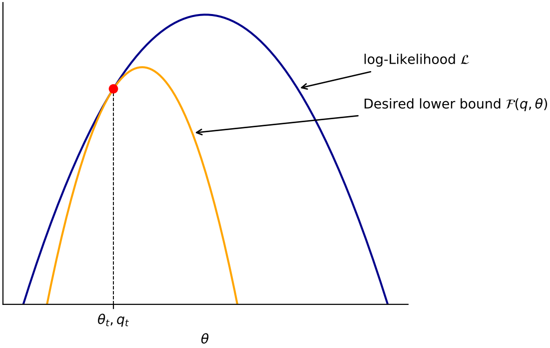
Kristian Kersting

TU Darmstadt

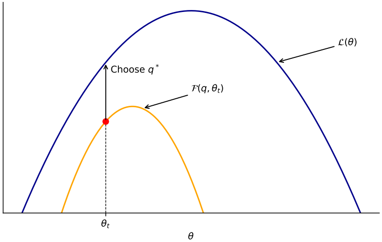
Summer Term 2020

Expectation Maximization

- Basic Idea



Expectation Maximization



Requirements

1. Guarantee a lower bound

$$\mathcal{F}(q, \theta) \leq \mathcal{L}(\theta) \quad \forall q, \theta$$

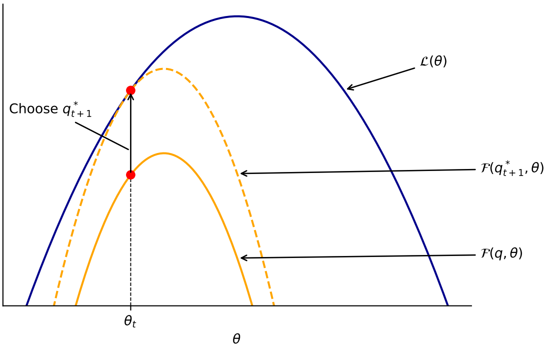
where q is the “guessed” distribution (aka surrogate function) and θ are the parameters

2. Choose q^* such that they touch

$$\mathcal{F}(q^*, \theta_t) = \mathcal{L}(\theta_t)$$

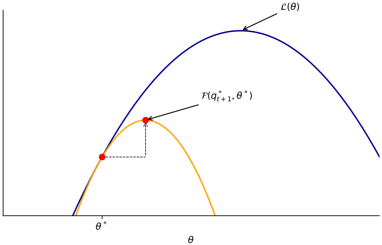
Expectation Maximization

- **Expectation-Step (E-Step)**



Expectation Maximization

- **Maximization-Step (M-Step)**



- Find θ^* by maximization

$$\theta^* = \arg\max_{\theta} \mathcal{F}\left(q^*_{t+1}, \theta\right)$$

Expectation Maximization

- Find a lower bound on $\mathcal{L}(\theta)$

$$
\begin{aligned}
\mathcal{L}(\theta) &= \sum_i \log p_\theta(x_i) \\
&= \sum_i \log \int p_\theta(x_i, z_i)\, \mathrm{d}z_i \\
&= \sum_i \log \int q(z_i) \frac{p_\theta(x_i, z_i)}{q(z_i)} \mathrm{d}z_i \\
& \text{(by Jensen's inequality)} \\
&\geq \sum_i \int q(z_i) \log \frac{p_\theta(x_i, z_i)}{q(z_i)} \mathrm{d}z_i \equiv \mathcal{F}(q, \theta) \\
\text{s.t.} \quad & \int q(z_i)\, \mathrm{d}z_i = 1 \quad \forall i
\end{aligned}
$$

Expectation Maximization

- Constrained Optimization Problem

$$\begin{aligned} \max \quad & \sum_i \int q(z_i) \log \frac{p_\theta(x_i, z_i)}{q(z_i)} \mathrm{d}z_i \\ \text{s.t.} \quad & \int q(z_i)\, \mathrm{d}z_i = 1 \quad \forall i \end{aligned}$$

Expectation Maximization

$$
\begin{aligned}
L &= \sum_i \left(\int q(z_i) \log \frac{p_\theta(x_i, z_i)}{q(z_i)} \mathrm{d}z_i \right) + \lambda_i \left(\int q_i(z_i) \mathrm{d}z_i - 1 \right) \quad \forall i \\
\nabla_{q(z_i)} L &= \left(\log \frac{p_\theta(x_i, z_i)}{q(z_i)} - 1 \right) + \lambda_i \stackrel{!}{=} 0 \\
&\Longrightarrow \quad q(z_i) = \exp(\lambda_i - 1)\, p_\theta(x_i, z_i) \\
\nabla_{\lambda_i} L &= \int q(z_i) \mathrm{d}z_i - 1 = 0 \\
& \exp(\lambda_i - 1) \int p_\theta(x_i, z_i) \mathrm{d}z_i = 0 \\
&\Longrightarrow \quad \lambda_i = 1 - \log \int p_\theta(x_i, z_i) \mathrm{d}z_i \\
& q(z_i) = p_\theta(z_i | x_i) \equiv \textbf{E-step}
\end{aligned}
$$

Expectation Maximization

1. We have a lower bound for the likelihood
2. We guaranteed
$$\mathcal{F}\left(q^*, \theta_t\right) = \mathcal{L}\left(\theta_t\right)$$
3. We want to guarantee
$$\mathcal{L}\left(\theta_{t+1}\right) \geq \mathcal{L}\left(\theta_t\right)$$
thus
$$\mathcal{L}\left(\theta_{t+1}\right) \geq \mathcal{F}\left(q^*_{t+1}, \theta_{t+1}\right) = \max_\theta \mathcal{F}\left(q^*_{t+1}, \theta\right) \geq \mathcal{L}\left(\theta_t\right)$$
4. Choose θ_{t+1} as
$$\theta_{t+1} = \arg\max_\theta \mathcal{F}\left(q^*_{t+1}, \theta\right)$$