Sheet 2

May 5, 2020

Exercise 1

a) Find optimal parameter λ_0

$$q_G(g) = \lambda * \frac{(g*\lambda)^a}{a!} e^{-g\lambda} \mathcal{H}(g), \lambda > 0, a \in \mathbb{N}_0$$

$$\mathcal{L}(\lambda) =< \ln q_G > \approx \frac{1}{K} \sum_{k=1}^{K} \ln q_G(\gamma^{(k)})$$

for K data samples $\gamma^{(1)}, \gamma^{(2)}, ..., \gamma^{(K)} \in \mathbb{R}_0^+$

1. Put together equations

$$\mathcal{L}(\lambda) = \frac{1}{K} \sum_{k=1}^{K} \ln(\lambda * \frac{(\gamma^{(k)} * \lambda)^a}{a!} e^{-\gamma^{(k)}\lambda} \mathcal{H}(\gamma^{(k)}))$$

2. Since all $\gamma^{(k)} \geq 0$: $\mathcal{H}(\gamma^{(k)}) = 1$

$$\mathcal{L}(\lambda) = \frac{1}{K} \sum_{k=1}^{K} \ln(\lambda * \frac{(\gamma^{(k)} * \lambda)^a}{a!} e^{-\gamma^{(k)}\lambda})$$

3. Use ln:

$$\mathcal{L}(\lambda) = \frac{1}{K} \sum_{k=1}^{K} \ln(\lambda) + a * \ln(\gamma^{(k)} * \lambda) - \ln(a!) - \gamma^{(k)} * \lambda$$

$$\mathcal{L}(\lambda) = \frac{1}{K} \sum_{k=1}^{K} \ln(\lambda) + a * \ln(\gamma^{(k)}) + a * \ln(\lambda) - \ln(a!) - \gamma^{(k)} * \lambda$$

4. Derive:

$$\frac{d}{d\lambda}\mathcal{L}(\lambda) = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{\lambda} + \frac{a}{\lambda} - \gamma^{(k)}$$

$$\frac{d}{d\lambda}\mathcal{L}(\lambda) = \frac{1}{K}\sum_{k=1}^{K}\frac{1+a}{\lambda} - \gamma^{(k)}$$

$$\frac{d}{d\lambda}\mathcal{L}(\lambda) = \frac{1+a}{\lambda} + \frac{1}{K}\sum_{k=1}^{K} -\gamma^{(k)}$$

5. Set to 0:

$$0 = \frac{1+a}{\lambda} + \frac{1}{K}\sum_{k=1}^{K} -\gamma^{(k)}$$

$$-\frac{1+a}{\lambda} = \frac{1}{K}\sum_{k=1}^{K} -\gamma^{(k)}$$

$$\frac{1+a}{\lambda} = -\frac{1}{K}\sum_{k=1}^{K} -\gamma^{(k)}$$

$$\frac{\lambda}{1+a} = -K * \frac{1}{\sum_{k=1}^{K} -\gamma^{(k)}}$$

$$\lambda_0 = -K(1+a) * \frac{1}{\sum_{k=1}^{K} -\gamma^{(k)}}$$

b) Interpretation of L^2 regularization term

$$\mathcal{L}_R(\lambda) = \mathcal{L}(\lambda) - \frac{C}{2}\lambda^2$$

We try to give λ minimal value, so we subtract the term. We square λ, because the sign does not matter, only that the final parameter is close to 0. C is the weight of the regularization, if it is large, we give the regularizer a high importance and our λ will be small.
C is the tradeoff between having a precise fit and having a constraint on λ (allowing worse fits, but closer to real risk)

c) Find optimal parameter λ^* of the regularized problem

$$\mathcal{L}_R(\lambda) = \mathcal{L}(\lambda) - \frac{C}{2}\lambda^2$$

1. Derive:

$$\frac{d}{d\lambda}\mathcal{L}_R(\lambda) = \frac{d}{d\lambda}\mathcal{L}(\lambda) - \frac{d}{d\lambda}\frac{C}{2}\lambda^2$$

$$\frac{d}{d\lambda}\mathcal{L}_R(\lambda) = \frac{1+a}{\lambda} + \frac{1}{K}\sum_{k=1}^{K} -\gamma^{(k)} - C\lambda$$

2. Set to 0:

$$0 = \frac{1+a}{\lambda^*} + \frac{1}{K}\sum_{k=1}^{K} -\gamma^{(k)} - C\lambda^*$$

$$\frac{1+a}{\lambda^*} - C\lambda^* = \frac{1}{K}\sum_{k=1}^{K} \gamma^{(k)}$$

$$1 + a - C\lambda^{*2} = \lambda\frac{1}{K}\sum_{k=1}^{K} \gamma^{(k)}$$

$$0 = C\lambda^{*2} + \lambda^*\frac{1}{K}\sum_{k=1}^{K} \gamma^{(k)} - (1+a)$$

$$0 = \lambda^{*2} + \lambda^*\frac{1}{KC}\sum_{k=1}^{K} \gamma^{(k)} - \frac{(1+a)}{C}$$

$$\lambda_1^*, \lambda_2^* = -\frac{1}{2KC}\sum_{k=1}^{K} \gamma^{(k)} \pm \sqrt{(\frac{1}{2KC}\sum_{k=1}^{K} \gamma^{(k)})^2 + \frac{(1+a)}{C}}$$

$\rightarrow$ only λ^* with + in front of the square root is valid, since the distribution has to be normalized, so λ^* has to be positive