Sheet 2

May 4, 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:

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

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

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

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

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

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

3. Check 2nd derivative:

$$\frac{d}{d\lambda^2}\mathcal{L}_R(\lambda) = \frac{1+a}{-\lambda^2} - C$$

$$\frac{d}{d\lambda^2}\mathcal{L}_R(\lambda) = \frac{1+a}{-(-\frac{1}{2KC}\sum_{k=1}^{K}\gamma^{(k)} \pm \sqrt{(\frac{1}{2KC}\sum_{k=1}^{K}\gamma^{(k)})^2 + \frac{(1+a)}{C}})^2} - C$$

case -:

$$\frac{d}{d\lambda^2}\mathcal{L}_R(\lambda_1) = \frac{1+a}{-((-\frac{1}{2KC}\sum_{k=1}^{K}\gamma^{(k)})^2 - (\frac{1}{2KC}\sum_{k=1}^{K}\gamma^{(k)})^2 + \frac{(1+a)}{C})} - C$$

$$\frac{d}{d\lambda^2}\mathcal{L}_R(\lambda_1) = \frac{1+a}{-(-\frac{1}{2KC}\sum_{k=1}^{K}\gamma^{(k)})^2 + (\frac{1}{2KC}\sum_{k=1}^{K}\gamma^{(k)})^2 - \frac{(1+a)}{C}} - C$$

$$\frac{d}{d\lambda^2}\mathcal{L}_R(\lambda_1) = \frac{1+a}{-\frac{(1+a)}{C}} - C$$

$$\frac{d}{d\lambda^2}\mathcal{L}_R(\lambda_1) = \frac{(1+a)C}{-(1+a)} - C$$

$$\frac{d}{d\lambda^2}\mathcal{L}_R(\lambda_1) = -C - C = -2C$$

case +:

$$\frac{d}{d\lambda^2}\mathcal{L}_R(\lambda_2) = \frac{1+a}{-((-\frac{1}{2KC}\sum_{k=1}^{K}\gamma^{(k)})^2 + (\frac{1}{2KC}\sum_{k=1}^{K}\gamma^{(k)})^2 + \frac{(1+a)}{C})} - C$$

$$\frac{d}{d\lambda^2}\mathcal{L}_R(\lambda_2) = \frac{1+a}{-(-\frac{1}{2KC}\sum_{k=1}^{K}\gamma^{(k)})^2 - (\frac{1}{2KC}\sum_{k=1}^{K}\gamma^{(k)})^2 - \frac{(1+a)}{C}} - C$$

$$\frac{d}{d\lambda^2}\mathcal{L}_R(\lambda_2) = -\frac{1+a}{2 * (\frac{1}{2KC}\sum_{k=1}^{K}\gamma^{(k)})^2 + \frac{(1+a)}{C}} - C$$

$$\frac{d}{d\lambda^2}\mathcal{L}_R(\lambda_2) = -\frac{2C^2(\sum_{k=1}^{K}\gamma^{(k)})^2}{a + 2C(\sum_{k=1}^{K}\gamma^{(k)})^2 + 1)}$$

$\rightarrow$ both 2nd derivatives are negative if C is positive

d) Find polynomial dependence of λ^* on C for small C using Taylor expansion

1. Original function:

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

2. Taylor expansion for C = 0 to second order: for $p = \frac{1}{K}\sum_{k=1}^{K}\gamma^{(k)}$ and $q = 1 + a$

Input interpretation:

series	$-\frac{p}{2\,C} - \sqrt{\left(\frac{p}{2\,C}\right)^2 - \frac{q}{C}}$	point	$C = 0$
		order	C^2

Series expansion at C = 0:

$$-\frac{p}{2\,C} - \frac{1}{2}\sqrt{\frac{p^2}{C^2}} + \frac{C\,q\sqrt{\frac{p^2}{C^2}}}{p^2} + O(C^2)$$

(generalized Puiseux series)

$\rightarrow$ correction term positive, since we add the first order derivative

Input interpretation:

series	$-\frac{p}{2C}+\sqrt{\left(\frac{p}{2C}\right)^2-\frac{q}{C}}$	point	$C = 0$
		order	C^2

Series expansion at C = 0:

$$-\frac{p}{2\,C}+\frac{1}{2}\sqrt{\frac{p^2}{C^2}}-\frac{C\,q\sqrt{\frac{p^2}{C^2}}}{p^2}+O(C^2)$$

(generalized Puiseux series)

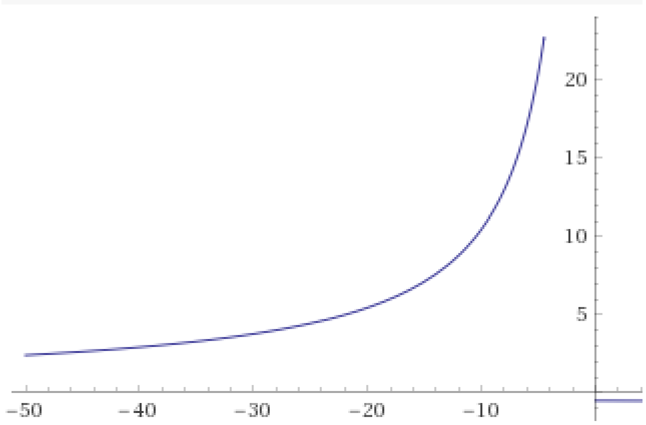
→ correction term negative, since we subtract the first order derivative

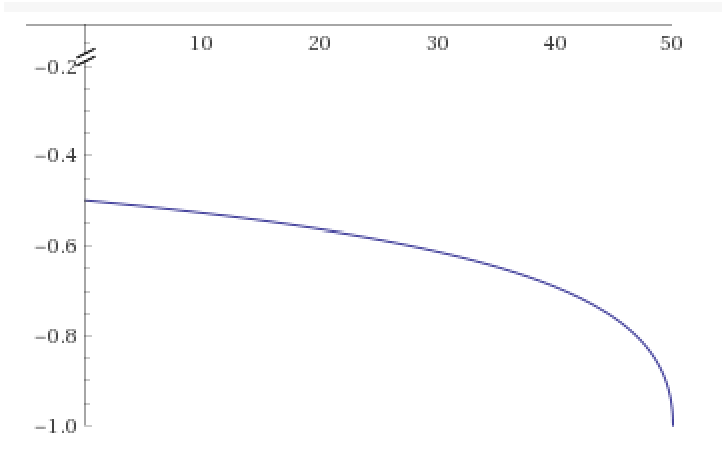
→ square root dependence in C

e) Scaling behaviour for $C \to \inf$ and sketch λ^* as a function of C

- scaling behaviour: if $C \to \infty$, $\lambda^* \to 0$. This matches what we said in the lecture: if we give the regularizer a high importance, the parameter will be small.
- plots (C on x-axis, $\lambda(C)$ on y-axis:

plot	$-\frac{p}{2C}+\sqrt{\left(\frac{p}{2C}\right)^2-\frac{q}{C}}$ where $p = 100, q = 50$





10
20
30
40
50
−0.2
−0.4
−0.6
−0.8
−1.0