Machine Learning for EDS

Tutorial week 1: Solutions

2023/2024

Important: this file is meant for students of the course Machine Learning for EDS (2023/2024) and is not allowed to be distributed to others.

Problem 1

1. For X a real-valued non-negative random variable,

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X\mathbb{1}_{X\geq a} + X\mathbb{1}_{X<a}] \\ &\geq \mathbb{E}[a\mathbb{1}_{X\geq a}] \\ &= a\mathbb{P}(X \geq a),\end{aligned}$$

where the inequality holds because $X\mathbb{1}_{X\geq a} \geq a\mathbb{1}_{X\geq a}$, since whenever the indicator is equal to one, we have $X \geq a$, and the second term in the sum is non-negative. Furthermore, we use that the expectation of an indicator function is equal to the probability of the condition of the indicator function being true, so $\mathbb{E}[\mathbb{1}_{X\geq a}] = \mathbb{P}(X \geq a)$. Thus:

$$\boxed{\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}}$$

2. (a) For any strictly increasing function $g : \mathbb{R} \to \mathbb{R}$, it holds that: $\mathbb{P}(X \geq a) = \mathbb{P}\Big(g(X) \geq g(a)\Big)$. Using the Markov inequality from 1.1 and using that X is b-sub-Gaussian, we thus have for any $s > 0$:

$$\mathbb{P}(X \geq a) = \mathbb{P}(e^{sX} \geq e^{sa}) \leq \frac{\mathbb{E}[e^{sX}]}{e^{sa}} \leq e^{\frac{s^2b^2}{2} - sa}.$$

$$\boxed{\forall s > 0\,, \quad \mathbb{P}(X \geq a) \leq e^{\frac{s^2b^2}{2} - sa}}$$

   (b) Taking the minimum over $s > 0$ on both sides of the inequality found in the previous question gives

$$\mathbb{P}(X \geq a) \leq \min_{s>0} \exp\left(\frac{s^2b^2}{2} - sa\right)$$

Because $\exp(\cdot)$ is increasing, it is clear that

$$\arg\min_{s>0} \exp\left(\frac{s^2b^2}{2} - sa\right) = \arg\min_{s>0} \frac{s^2b^2}{2} - sa$$

Now take first and second order derivatives of $f(s) = \frac{s^2b^2}{2} - sa$, find the critical point and check that the function is strictly convex by verifying that the second order derivative is always strictly positive. Then you can conclude that the critical point $s^* = a/b^2$ must be the unique global minimiser of f. Filling in this value yields the required result:

$$\mathbb{P}(X \geq a) \leq \min_{s>0} e^{\frac{s^2b^2}{2} - sa} = e^{\frac{(s^*)^2b^2}{2} - s^*a} = \exp\Big(-\frac{a^2}{2b^2}\Big).$$

$$\boxed{\mathbb{P}(X \geq a) \leq \exp\Big(-\frac{a^2}{2b^2}\Big)}$$

Problem 2

1. (a) Because $X_1, \dots, X_n$ are assumed independent and b_i-sub-Gaussian, we have that for any $t \in \mathbb{R}$:

$$\mathbb{E}[e^{tS_n}] = \mathbb{E}[e^{t\sum_{i=1}^n X_i}] = \prod_{i=1}^n \mathbb{E}[e^{tX_i}] \leq \prod_{i=1}^n \exp\Big(\frac{t^2b_i^2}{2}\Big) = \exp\Big(\frac{t^2\sum_{i=1}^n b_i^2}{2}\Big).$$

This shows that S_n is itself b-sub-Gaussian with $b = \sqrt{\sum_{i=1}^n b_i^2}$.

(b) As S_n is b-sub-Gaussian with $b = \sqrt{\sum_{i=1}^n b_i^2}$, we can apply the result of point 2 of problem 1, to get that for any $a > 0$:

$$\forall a > 0, \quad \mathbb{P}(S_n \geq a) \leq \exp\Big(-\frac{a^2}{2\sum_{i=1}^n b_i^2}\Big).$$

Furthermore, if X is b-sub-Gaussian, then $-X$ is also automatically b-sub-Gaussian since the definition should hold for any $s \in \mathbb{R}$. Hence, it follows that $-S_n$ is b-sub-Gaussian, so

$$\forall a > 0, \quad \mathbb{P}(S_n \leq -a) = \mathbb{P}(-S_n \geq a) \leq \exp\Big(-\frac{a^2}{2\sum_{i=1}^n b_i^2}\Big).$$

In conclusion:

$$\boxed{\begin{aligned} \forall a > 0, \quad \mathbb{P}(S_n \geq a) \leq \exp\Big(-\frac{a^2}{2\sum_{i=1}^n b_i^2}\Big). \\ \forall a > 0, \quad \mathbb{P}(S_n \leq -a) \leq \exp\Big(-\frac{a^2}{2\sum_{i=1}^n b_i^2}\Big). \end{aligned}}$$

(c) Using the hint we have $\mathbb{P}(|S_n| \geq a) = \mathbb{P}(\{S_n \geq a\} \cup \{S_n \leq -a\}) = \mathbb{P}(S_n \geq a) + \mathbb{P}(S_n \leq -a)$, where the final equality holds because $\{S_n \geq a\}$ and $\{S_n \leq -a\}$ are disjoint events, so σ-additivity applies. From the result in (b) it now follows that:

$$\boxed{\forall a > 0, \quad \mathbb{P}(|S_n| \geq a) = \mathbb{P}(S_n \geq a) + \mathbb{P}(S_n \leq -a) \leq 2\exp\Big(-\frac{a^2}{2\sum_{i=1}^n b_i^2}\Big).}$$

2. (a) Let us start with a standard Gaussian random variable: $X \sim \mathcal{N}(0,1)$. We have

$$\mathbb{E}[e^{tX}] = \int_{\mathbb{R}} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = e^{\frac{t^2}{2}} \underbrace{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}} dx}_{=1} = e^{\frac{t^2}{2}}.$$

Thus, X in particular satisfies $\mathbb{E}[e^{tX}] \leq e^{\frac{t^2}{2}}$ for all $t \in \mathbb{R}$ and is 1-sub-Gaussian. Consider now $\tilde{X} = \mathcal{N}(\mu, \sigma^2)$. It holds that $\mathbb{E}[X] = \mu$ and $\tilde{X} \overset{d}{=} \sigma X + \mu$. Hence, for any $t \in \mathbb{R}$

$$\mathbb{E}\Big[\exp\Big(t(\tilde{X} - \mathbb{E}[\tilde{X}])\Big)\Big] = \mathbb{E}\Big[\exp\Big(\sigma t X\Big)\Big] = e^{(\sigma t)^2/2},$$

which shows that $\tilde{X} - \mathbb{E}[\tilde{X}]$ is σ-sub-Gaussian.

(b) Consider now $X_1, \ldots, X_n$ be n independent random variables such that for any i, $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$. From point 1 and using the fact that the X_i's are independent and σ_i-sub-Gaussian, we have: for any $a > 0$:

$$\boxed{\begin{array}{c}\text{For } S_n = \sum_{i=1}^{n} \Big(X_i - \mathbb{E}[X_i]\Big), \\ \mathbb{P}(S_n \geq a) \leq \exp\Big(-\frac{a^2}{2\sum_{i=1}^n \sigma_i^2}\Big).\end{array}}$$

Problem 3

1. Notice that

$$\frac{d-x}{d-c}c + \frac{x-c}{d-c}d = x.$$

Therefore, we can write

$$e^{tx} = \exp\left(\frac{d-x}{d-c}tc + \frac{x-c}{d-c}td\right)$$

Now also notice that $\frac{d-x}{d-c}$ takes values in $[0,1]$ because x takes values in $[c,d]$ with $c < d$, and furthermore $\frac{d-x}{d-c} = 1 - \frac{x-c}{d-c}$. Notice also that the function $f(x) = \exp(x)$ is convex (check it yourself!). Hence: using the convexity of the exponential function, we have that for any $x \in [c,d]$ and $t \in \mathbb{R}$,

$$e^{tx} = \exp\left(\frac{d-x}{d-c}tc + \frac{x-c}{d-c}td\right) \leq \frac{d-x}{d-c}e^{tc} + \frac{x-c}{d-c}e^{td},$$

2. It follows from the inequality derived in point one that:

$$\mathbb{E}[e^{tX}] \leq \frac{d-\mathbb{E}[X]}{d-c}e^{tc} + \frac{\mathbb{E}[X]-c}{d-c}e^{td} = \frac{d}{d-c}e^{tc} + \frac{-c}{d-c}e^{td},$$

because it is given in the question that $\mathbb{E}[X] = 0$

3. Letting $h := t(d-c)$, $p := \dfrac{-c}{d-c}$ and $L : h \mapsto L(h) := -hp + \ln(1-p+pe^h)$, we have that:

$$\begin{aligned} e^{L(h)} &= \exp(-hp + \ln(1-p+pe^h)) \\ &= \exp(-hp)(1-p+pe^h) \\ &= \exp(tc)\left(\frac{d}{d-c} + \frac{-c}{d-c}\exp(t(d-c))\right) \\ &= \frac{d}{d-c}\, e^{tc} + \frac{-c}{d-c}\, e^{td}. \end{aligned}$$

Combining with point 2 it follows that $\mathbb{E}[e^{tX}] \le e^{L(h)}$.

4. It is clear that $L(0) = 0 + \ln(1-p+p) = 0$. Furthermore we have:

$$\begin{aligned} L'(h) &= -p + \frac{pe^h}{1-p+pe^h}, \\ L''(h) &= \frac{pe^h(1-p)}{(1-p+pe^h)^2}. \end{aligned}$$

Thus: $L'(0) = -p + p = 0$, and using the general inequality $a^2 + b^2 \ge 2ab$ (which follows from $(a-b)^2 \ge 0$), we have for any h:
$L''(h) = \dfrac{pe^h(1-p)}{(1-p)^2 + 2pe^h(1-p) + (pe^h)^2} \le \dfrac{pe^h(1-p)}{4pe^h(1-p)} = 1/4$.

5. Using a second order Taylor expansion of $h \mapsto L(h)$ around $h = 0$, we know there exists an h_0 between 0 and h such that

$$L(h) = L(0) + hL'(0) + \frac{h^2}{2}L''(h_0) = \frac{h^2}{2}L''(h_0) \le \frac{h^2}{8}$$

6. Therefore $L(h) \le t^2(d-c)^2/8$. Hence, using the result in point 3, for any $t \in \mathbb{R}$,

$$\mathbb{E}[e^{tX}] \le e^{L(h)} \le \exp\left(\frac{t^2(d-c)^2}{8}\right),$$

which establishes that X is sub-Gaussian with parameter $\dfrac{d-c}{2}$ (notice that the second inequality is valid because $\exp(\cdot)$ is an increasing function).

7. In that case consider $\tilde{X} = X - \mathbb{E}[X]$, which is centered and bounded between $[c - \mathbb{E}[X], d - \mathbb{E}[X]]$. Then it can be shown in the same way as above that $X - \mathbb{E}X$ is b- sub-Gaussian with $b = (d-c)/2$.

8. Letting now $X_1, \ldots, X_n$ be n independent bounded random variables such that for any i, $X_i \in [c_i, d_i]$. Then, for any i, $X_i - \mathbb{E}[X_i]$ is sub-Gaussian with parameter $\dfrac{d_i - c_i}{2}$, and by point 1 of Problem 2, we obtain

$$\boxed{\begin{aligned}&\text{For any } a>0,\ \text{with } S_n=\sum_{i=1}^{n}\Big(X_i-\mathbb{E}[X_i]\Big),\\ &\qquad\qquad \mathbb{P}(S_n\geq a)\leq \exp\left(-\frac{2a^2}{\sum_{i=1}^n (d_i-c_i)^2}\right)\\ &\qquad\qquad \mathbb{P}(S_n\leq -a)\leq \exp\left(-\frac{2a^2}{\sum_{i=1}^n (d_i-c_i)^2}\right)\\ &\text{and}\qquad \mathbb{P}(|S_n|\geq a)\leq 2\exp\left(-\frac{2a^2}{\sum_{i=1}^n (d_i-c_i)^2}\right)\end{aligned}}$$

Problem 4

1. Using that the function $\exp(sx)$ is convex in x for any $s>0$, we have by Jensen's inequality that

$$e^{s\mathbb{E}[M]}\leq \mathbb{E}\left[e^{sM}\right]=\mathbb{E}\left[e^{s\max\limits_{1\leq k\leq K} Z_k}\right]=\mathbb{E}\left[\max_{1\leq k\leq K} e^{sZ_k}\right]\leq \mathbb{E}\left[\sum_{k=1}^{K} e^{sZ_k}\right]=\sum_{k=1}^{K}\mathbb{E}\left[e^{sZ_k}\right]$$

2. Starting from the result from point 1, we have for any $s>0$:

$$e^{s\mathbb{E}[M]}\leq \sum_{k=1}^{K}\mathbb{E}\left[e^{sZ_k}\right]\leq \sum_{k=1}^{K} e^{s^2\nu^2/2}=Ke^{s^2\nu^2/2}\,,$$

where for the second inequality we use that each Z_k is ν-sub-Gaussian.

3. Starting from the result from point 2 and by taking the natural logarithm and dividing by s on both sides of the inequality (note: $\ln(\cdot)/s$ is an increasing function so we do not have to flip the sign), we get for each $s>0$

$$\mathbb{E}[M]\leq \frac{1}{s}\ln\left(Ke^{\frac{s^2\nu^2}{2}}\right)=\frac{\ln(K)}{s}+\frac{s\nu^2}{2}\,.$$

Hence, it follows that

$$\mathbb{E}[M]\leq \inf_{s>0}\left\{\frac{\ln(K)}{s}+\frac{s\nu^2}{2}\right\}$$

4. We want to find the infimum on the right hand side of the inequality above. Taking the first and second order derivative gives:

$$\begin{aligned}\frac{\partial}{\partial s}\left(\frac{\ln(K)}{s}+\frac{s\nu^2}{2}\right)&=-\frac{\ln(K)}{s^2}+\frac{\nu^2}{2}\,,\\ \frac{\partial^2}{\partial s^2}\left(\frac{\ln(K)}{s}+\frac{s\nu^2}{2}\right)&=\frac{2\ln(K)}{s^3}\,.\end{aligned}$$

The second order derivative is strictly positive for every $s > 0$ and any $K > 1$, meaning that the function is strictly convex, which implies that the critical point of the function is its minimum. The critical point is:

$$-\frac{\ln(K)}{s^2} + \frac{\nu^2}{2} = 0 \implies s^2 = 2\frac{\ln(K)}{\nu^2} \implies s = \frac{1}{\nu}\sqrt{2\ln(K)}$$

(Notice that $s = -\sqrt{2\ln(K)}/\nu$ is not an option, because we need $s > 0$). Plugging this value of s into the expression we obtain:

$$\mathbb{E}[M] \leq \inf_{s>0}\left\{\frac{\ln(K)}{s} + \frac{s\nu^2}{2}\right\} = \frac{\nu\ln(K)}{\sqrt{2\ln(K)}} + \frac{\sqrt{2\ln(K)}\nu^2}{2\nu} = \nu\sqrt{2\ln(K)}\,.$$