Exercises on Chapter 3: Statistics and concentration inequalities

Mathematics of Data Science, M1 IDD

November-December 2023*

Exercise 3.1: Boosting

Suppose that we perform $2\,m$ independent runs of a randomized algorithm designed to solve a decision problem (e.g. is a given convex optimization problem feasible?). Because of the randomness, the algorithm is only correct with probability $\frac{1}{2}+\delta$ for some $\delta \in (0,1)$. To make a decision, we choose the output returned by the majority of runs.

a) Let y_i be a Bernoulli random variable such that $y_i = 1$ if the ith run returns the wrong output, and $y_i = 0$ otherwise. Compute $\mathbb{E}\left[y_i\right]$.

b) Express the probability of making the right conclusion from the output of the $2m$ instances.

c) Let $p \in [0,1)$. Using Hoeffding's inequality, show that the probability of making the right conclusion is at least $1-p$ when

$$m \geq \frac{1}{4\delta^2} \ln\left(\frac{1}{p}\right).$$

*Last updated December 20, 2023.

Exercise 3.2: Chernoff inequalities

In this exercise, we study another type of concentration inequalities than that seen in class called *Chernoff bounds* or *Chernoff inequalities*. In the general form, this inequality states that for any random variable Y and any $t \in \mathbb{R}$, we have

$$\mathbb{P}(Y \geq t) \quad \leq \quad \min_{\lambda \geq 0} \mathbb{E}\left[\exp(\lambda(Y - t))\right]. \tag{1}$$

a) Proving (1) amounts to proving

$$\ln\left(\mathbb{P}(Y \geq t)\right) \quad \leq \quad \min_{\lambda \geq 0} \ln\left(\mathbb{E}\left[\exp(\lambda(Y - t))\right]\right). \tag{2}$$

Justify that right-hand side of (2) is the solution to a convex optimization problem. *To this end, you may use a generalization of the Hölder inequality from Exercise 1.8, that states that for any random variables w, z, we have*

$$\mathbb{E}_{w,z}\left[w\, z\right] \leq \mathbb{E}_w\left[|w|^p\right]^{1/p} \mathbb{E}_z\left[|z|^q\right]^{1/q}$$

any pair (p, q) such that $p > 1, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$.

b) Suppose that $y \sim \mathcal{N}(0, 1)$. In that case, one can show that $\ln\left(\mathbb{E}\left[\exp(\lambda y)\right]\right) = \frac{\lambda^2}{2}$. Use this property to deduce from (1) that

$$\mathbb{P}(Y \geq t) \quad \leq \quad \exp\left(-\frac{t^2}{2}\right).$$

Exercise 3.3: Chernoff inequalities for vectors

In this exercise, we seek a Chernoff-type bound in a vector setting. More precisely, we consider a Gaussian vector $\boldsymbol{y} \sim \mathcal{N}(\mathbf{0}_{\mathbb{R}^n}, \boldsymbol{I}_n)$ and a nonempty polyhedral set defined by $\mathcal{C} = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}\}$ with $\boldsymbol{A} \in \mathbb{R}^{\ell \times n}$ and $\boldsymbol{b} \in \mathbb{R}^\ell$. Our goal is to provide a bound of the form

$$\mathbb{P}(\boldsymbol{y} \in \mathcal{C}) \le \mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y} + \mu\right)\right] \tag{3}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^n$ and $\mu \in \mathbb{R}$. As in the previous exercise, we would like to obtain the tightest bound possible.

a) Using that $\mathbb{P}(\boldsymbol{y} \in \mathcal{C}) = \mathbb{E}[1_{\mathcal{C}}(\boldsymbol{y})]$, justify that any pair $(\boldsymbol{\lambda}, \mu) \in \mathbb{R}^n \times \mathbb{R}$ satisfying $\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y} + \mu\right) \ge 1_{\mathcal{C}}(\boldsymbol{y})$ for every $\boldsymbol{y} \in \mathbb{R}^n$ also satisfies (3) with $-\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y} \le \mu \; \forall \boldsymbol{y} \in \mathcal{C}$.

b) By considering logarithms, show that

$$\ln\left(\mathbb{P}(\boldsymbol{y} \in \mathcal{C})\right) \le \min_{\boldsymbol{\lambda} \in \mathbb{R}^n} \left\{ S_{\mathcal{C}}(-\boldsymbol{\lambda}) + \ln \mathbb{E}\left[e^{\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{z}}\right] \right\}, \tag{4}$$

with $S_{\mathcal{C}} : \boldsymbol{y} \mapsto \max_{\boldsymbol{x} \in \mathcal{C}} \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}$.

c) Since $\boldsymbol{y}$ is Gaussian, we have that $\ln(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}\right)\right]) = \frac{\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\lambda}}{2}$ for any $\boldsymbol{\lambda}$. In addition, we can show that

$$S_{\mathcal{C}}(\boldsymbol{y}) = \min_{\boldsymbol{u} \in \mathbb{R}^\ell} \{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{u} | \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} = \boldsymbol{y}, \boldsymbol{u} \ge \mathbf{0}\}$$

for any $\boldsymbol{y} \in \mathbb{R}^n$. Show then that the right-hand side of (4) corresponds to the optimal value of the quadratic problem

$$\begin{array}{lll} \text{minimize}_{\boldsymbol{\lambda} \in \mathbb{R}^n, \boldsymbol{v} \in \mathbb{R}^\ell} & \boldsymbol{b}^{\mathrm{T}}\boldsymbol{v} + \frac{\|\boldsymbol{\lambda}\|^2}{2} \\ \text{s.t.} & \boldsymbol{v} \ge \mathbf{0}, \\ & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{v} + \boldsymbol{\lambda} = \mathbf{0}. \end{array} \tag{5}$$

d) The problem (5) is equivalent to

$$\underset{\boldsymbol{v} \in \mathbb{R}^\ell}{\text{minimize}}\, \boldsymbol{b}^{\mathrm{T}}\boldsymbol{v} + \frac{\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{v}\|^2}{2} \quad \text{s.t.} \quad \boldsymbol{v} \ge \mathbf{0}, \tag{6}$$

where we reformulated the problem so as to eliminate the $\boldsymbol{\lambda}$ variables while preserving the same optimal value.

i) Using that same reformulation technique, show that the dual of problem (6) is equivalent to

$$\begin{array}{ll} \text{maximize}_{\boldsymbol{x} \in \mathbb{R}^m} & -\frac{\|\boldsymbol{x}\|^2}{2} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}. \end{array} \tag{7}$$

ii) Justify that the optimal value of problem (7) is $-\frac{1}{2}\mathrm{dist}(\mathbf{0}, \mathcal{C})^2$, where $\mathrm{dist}(\boldsymbol{a}, \mathcal{C}) = \min_{\boldsymbol{y} \in \mathcal{C}} \|\boldsymbol{y} - \boldsymbol{a}\|$.

iii) Strong duality holds for problem (6). Using this property, provide a closed-form expression for (4) and (3).

Solutions

Solution for Exercise 3.1: Boosting

Question a) A straightforward calculation gives

$$\mathbb{E}\left[Y_i\right] = 1 \times \mathbb{P}\left(Y_i = 1\right) + 0 \times \mathbb{P}\left(Y_i = 0\right) = 1 - \left(\frac{1}{2} + \delta\right) = \frac{1}{2} - \delta.$$

Question b) Since the decision is made on $2m$ runs of the algorithm by majority voting, we know that the correct output is accepted if $\sum_{i=1}^{2m} y_i < m$, since this is only possible when more than half of the voters returned 0. As a result, the probability of making the right decision is

$$\mathbb{P}\left(\sum_{i=1}^{2m} y_i < m\right).$$

Question c) We use the variant of Hoeffding's inequality tailored to bounded random variables. [1] For any $t \geq 0$, we have

$$\mathbb{P}\left(\sum_{i=1}^{2m} \left(y_i - \mathbb{E}\left[y_i\right]\right) \geq t\right) \leq \exp\left[-\frac{2t^2}{2m}\right] = \exp\left[-\frac{t^2}{m}\right],$$

where the $2m$ factor on the right-hand side corresponds to the squared norm of the vector of all ones. Using the formula for the expected value, we obtain

$$\mathbb{P}\left(\sum_{i=1}^{2m} y_i \geq t + m - 2m\delta\right) \leq \exp\left[-\frac{t^2}{m}\right].$$

Setting $t = 2m\delta > 0$ gives

$$\mathbb{P}\left(\sum_{i=1}^{2m} y_i \geq m\right) \leq \exp\left[-\frac{4m^2\delta^3}{m}\right] = \exp\left[-4m\delta^2\right]$$

Our goal is to guarantee that $\mathbb{P}\left(\sum_{i=1}^{2m} y_i < m\right) \geq 1 - p$, which is equivalent to $\mathbb{P}\left(\sum_{i=1}^{2m} y_i \geq m\right) < p$. Choosing $m > \frac{1}{2\delta^2} \ln\left(\frac{1}{p}\right)$, we see that

$$\exp\left[-2m\delta^2\right] < \exp\left[-\ln\left(\tfrac{1}{p}\right)\right] = p,$$

and the desired conclusion follows.

[1] In its general form, this inequality states that for any set of variables $y_1, \ldots, y_N$ that are bounded in $[m, M]$ and any $t \geq 0$, we have

$$\mathbb{P}\left(\sum_{i=1}^{N} \left(y_i - \mathbb{E}\left[y_i\right]\right) \geq t\right) \leq \exp\left[-\frac{2t^2}{N(M-m)^2}\right].$$

Solution for Exercise 3.2: Chernoff inequality

Foreword to question a) The equivalence between (1) and (2) can be justified as follows. Suppose that $\lambda \geq 0$ satisfies $\mathbb{P}(y \geq t) \leq \mathbb{E}[\exp(\lambda(y-t))]$ for all $t \in \mathbb{R}$. Then, by taking logarithms on both sides of the inequality (allowing $\ln(0) = -\infty$ and $-\infty \leq -\infty$), we obtain

$$\ln(\mathbb{P}(y \geq t)) \leq \ln(\mathbb{E}[\exp(\lambda(y-t))]).$$

It remains to show that minimizing the right-hand side of the latter inequality over λ gives the same bound that that obtained by (1). For any $\mu \geq 0$, using that the exponential function is monotonically increasing gives

$$\exp\left[\min_{\lambda \geq 0} \ln(\mathbb{E}[\exp(\lambda(y-t))])\right] \leq \exp[\ln(\mathbb{E}[\exp(\mu(y-t))])] = \mathbb{E}[\exp(\mu(y-t))].$$

Hence

$$\begin{aligned} \exp[\min_{\lambda \geq 0} \ln(\mathbb{E}[\exp(\lambda(y-t))])] &\leq \min_{\mu \geq 0} \mathbb{E}[\exp(\mu(y-t))] \\ \Rightarrow \qquad \min_{\lambda \geq 0} \ln(\mathbb{E}[\exp(\lambda(y-t))]) &\leq \ln[\min_{\mu \geq 0} \mathbb{E}[\exp(\mu(y-t))]]. \end{aligned}$$

Conversely, using that $\ln(\cdot)$ is monotonically increasing gives

$$\begin{aligned} \ln[\min_{\mu \geq 0} \mathbb{E}[\exp(\mu(y-t))]] &\leq \ln[\mathbb{E}[\exp(\lambda(y-t))]] \\ \Rightarrow \ln[\min_{\mu \geq 0} \mathbb{E}[\exp(\mu(y-t))]] &\leq \min_{\lambda \geq 0} \ln[\mathbb{E}[\exp(\lambda(y-t))]] \end{aligned}$$

Overall, we have shown that

$$\ln\left[\min_{\lambda \geq 0} \mathbb{E}[\exp(\mu(y-t))]\right] = \min_{\lambda \geq 0} \ln[\mathbb{E}[\exp(\lambda(y-t))]],$$

and therefore the two equalities are equivalent.

Question a) Observe that

$$\min_{\lambda \geq 0} \ln(\mathbb{E}[\exp(\lambda(y-t))]) \quad = \quad \min_{\lambda \geq 0} \ln(\mathbb{E}[\exp(\lambda y)]) - \lambda t.$$

Since $t \mapsto -\lambda t$ is a linear function of λ, it is convex, and thus it suffices to show that $\lambda \mapsto \ln(\mathbb{E}[\exp(\lambda y)])$ is convex on $\mathbb{R}_+$ to arrive at the desired result.

To this end, we consider two values $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$, as well as $\alpha \in [0,1]$. Our goal is to prove

$$\ln(\mathbb{E}[\exp((\alpha\lambda_1 + (1-\alpha)\lambda_2)y)]) \leq \alpha \ln(\mathbb{E}[\exp(\lambda_1 y)]) + (1-\alpha)\ln(\mathbb{E}[\exp(\lambda_2 y)]).$$

If $\alpha \in \{0,1\}$, the result trivially holds. Otherwise, we apply Minkowski's inequality to $\mathbb{E}[\exp((\alpha\lambda_1 + (1-\alpha)\lambda_2)y)] = \mathbb{E}[\exp(\alpha\lambda_1 y) \times \exp((1-\alpha)\lambda_2 y)]$. Using $Y = \alpha\lambda_1 y$, $Z = (1-\alpha)\lambda_2 y$, $p = \frac{1}{\alpha}$ and $q = \frac{1}{1-\alpha}$. We obtain

$$\begin{aligned} \mathbb{E}[\exp((\alpha\lambda_1 + (1-\alpha)\lambda_2)y)] &\leq \mathbb{E}\left[\exp(\alpha\lambda_1 y)^{1/\alpha}\right]^{\alpha} \mathbb{E}[\exp((1-\alpha)\lambda_2 y)^{\alpha}]^{1-\alpha} \\ &= \mathbb{E}[\exp(\lambda_1 y)]^{\alpha} \mathbb{E}[\exp(\lambda_2 y)]^{1-\alpha}, \end{aligned}$$

Taking logarithms then leads to

$$\ln(\mathbb{E}[\exp((\alpha\lambda_1 + (1-\alpha)\lambda_2)y)]) \leq \alpha \ln(\mathbb{E}[\exp(\lambda_1 y)]) + (1-\alpha)\ln(\mathbb{E}[\exp(\lambda_2 y)]),$$

showing that the function is indeed convex.

Question b) By applying $\ln \mathbb{E}[\exp(\lambda Y)] = \frac{\lambda^2}{2}$ in the inequality derived in question a), we obtain that

$$\ln[\mathbb{P}(Y \geq t)] \leq \min_{\lambda \geq 0} \left\{ -\lambda t + \frac{\lambda^2}{2} \right\}$$

The objective function of the right-hand side optimization problem is a convex quadratic in λ, and its minimum is attained at $\lambda^* = \max\{t, 0\}$. Indeed, if $t \geq 0$, then the minimum is $\lambda^* = t \geq 0$, while if $t < 0$, we have $-\lambda t + \frac{\lambda^2}{2} \geq 0$, hence $\lambda^* \geq 0$ is a minimum[2]. When $t \geq 0$, the inequality gives

$$\ln[\mathbb{P}(Y \geq t)]] \leq -\frac{t^2}{2},$$

and thus $\mathbb{P}(Y \geq t) \leq \exp(-t^2/2)$. Since a probability is always bounded above by 1 and $\exp(-t^2/2) > \exp(0) = 1$ for any $t < 0$, the inequality remains valid when $t < 0$, proving the desired result.

Solution for Exercise 3.3: Chernoff inequalities for vectors

Question a) Consider the function

$$\begin{array}{rccl} f: & \mathbb{R}^n & \to & \mathbb{R} \\ & \boldsymbol{y} & \mapsto & \exp(\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{y} + \mu). \end{array}$$

such that $f(\boldsymbol{y}) \geq 1_{\mathcal{C}}(\boldsymbol{y})$ for every $\boldsymbol{y} \in \mathbb{R}^n$. By definition of the indicator function, it implies that

$$\begin{aligned} f(\boldsymbol{y}) \geq 1_{\mathcal{C}}(\boldsymbol{y}) \ \forall \boldsymbol{y} \in \mathbb{R}^n \quad & \Leftrightarrow \quad \left\{ \begin{array}{ll} f(\boldsymbol{y}) \geq 1 & \forall \boldsymbol{y} \in \mathcal{C} \\ f(\boldsymbol{y}) \geq 0 & \forall \boldsymbol{y} \notin \mathcal{C} \end{array} \right. \\ & \Leftrightarrow \quad \left\{ \begin{array}{ll} \exp(\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{y} + \mu) \geq 1 & \forall \boldsymbol{y} \in \mathcal{C} \\ \exp(\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{y} + \mu)) \geq 0 & \forall \boldsymbol{y} \notin \mathcal{C} \end{array} \right. \\ & \Leftrightarrow \quad -\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{y} \leq \mu \quad \forall \boldsymbol{y} \in \mathcal{C}, \end{aligned}$$

where the latter equivalence comes from the fact that an exponential is always positive, hence the inequalities for $\boldsymbol{y} \notin \mathcal{C}$ always hold.

Question b) Taking logarithms on both sides of (3) gives

$$\ln(\mathbb{P}(\boldsymbol{z} \in \mathcal{C})) \leq \ln\left(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{y} + \mu\right)\right]\right).$$

We want to compute the pair $(\boldsymbol{\lambda}, \mu)$ that yields the tightest bound. From question a), we know that such a pair must satisfy $-\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{y} \leq \mu \ \forall \boldsymbol{y} \in \mathcal{C}$. As a result, the best lower bound is given as an optimal value of an optimization problem over $\boldsymbol{\lambda}$ and μ, namely

$$\min_{\boldsymbol{\lambda}, \mu} \left\{ \ln\left(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{y} + \mu\right)\right]\right) \middle| -\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{y} \leq \mu \ \forall \boldsymbol{y} \in \mathcal{C} \right\}.$$

Using $\ln\left(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{y} + \mu\right)\right]\right) = \ln\left(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{y}\right)\right]\right) + \mu$, the problem can be rewritten as

$$\min_{\boldsymbol{\lambda}, \mu} \left\{ \mu + \ln\left(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{y}\right)\right]\right) \middle| -\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{y} \leq \mu \ \forall \boldsymbol{y} \in \mathcal{C} \right\}.$$

[2] One can also establish this using the optimality conditions from Chapter 2.

Now, since the objective function minimizes μ and $\mu \geq -\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}\ \forall \boldsymbol{y} \in \mathcal{C}$, the optimal μ for a given $\boldsymbol{\lambda}$ is $\max_{\boldsymbol{y}\in\mathcal{C}} -\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}$. As a result, we can reformulate the problem as a problem involving only $\boldsymbol{\lambda}$:

$$\begin{aligned}
& \min_{\boldsymbol{\lambda},\mu} \left\{\mu + \ln\left(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}\right)\right]\right) \middle| -\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y} \leq \mu\ \forall \boldsymbol{y} \in \mathcal{C}\right\} \\
= & \min_{\boldsymbol{\lambda}} \min_{\mu} \left\{\mu + \ln\left(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}\right)\right]\right) \middle| -\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y} \leq \mu\ \forall \boldsymbol{y} \in \mathcal{C}\right\} \\
= & \min_{\boldsymbol{\lambda}} \left\{\max_{\boldsymbol{x}\in\mathcal{C}}[-\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{x}] + \ln(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}\right)\right])\right\} \\
= & \min_{\boldsymbol{\lambda}} \left\{S_{\mathcal{C}}(-\boldsymbol{\lambda}) + \ln(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}\right)\right])\right\}.
\end{aligned}$$

As a result, we must have

$$\ln\left(\mathbb{P}\left(\boldsymbol{y} \in \mathcal{C}\right)\right) \leq \min_{\boldsymbol{\lambda}} \left\{S_{\mathcal{C}}(-\boldsymbol{\lambda}) + \ln(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}\right)\right])\right\}.$$

Question c) Using the property of the Gaussian vector $\boldsymbol{y}$, we have

$$\ln\left(\mathbb{P}\left(\boldsymbol{y} \in \mathcal{C}\right)\right) \leq \min_{\boldsymbol{\lambda}} \left\{S_{\mathcal{C}}(-\boldsymbol{\lambda}) + \frac{\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\lambda}}{2}\right\}.$$

Combining this with the result of question b) gives

$$\begin{aligned}
\min_{\boldsymbol{\lambda}\in\mathbb{R}^m} \left\{S_{\mathcal{C}}(-\boldsymbol{\lambda}) + \tfrac{\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\lambda}}{2}\right\} &= \min_{\boldsymbol{\lambda}\in\mathbb{R}^m} \left\{\min_{\boldsymbol{u}\in\mathbb{R}^\ell} \left\{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{u} \mid \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} + \boldsymbol{\lambda} = \boldsymbol{0}, \boldsymbol{u} \geq \boldsymbol{0}\right\} + \tfrac{\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\lambda}}{2}\right\} \\
&= \min_{\substack{\boldsymbol{\lambda}\in\mathbb{R}^m \\ \mu\in\mathbb{R}^\ell}} \left\{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{u} + \tfrac{\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\lambda}}{2} \middle| \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} + \boldsymbol{\lambda} = \boldsymbol{0}, \boldsymbol{u} \geq \boldsymbol{0}\right\},
\end{aligned}$$

which is the desired result.

d)i) The problem (6) is convex and its feasible set has a nonempty interior. Slater's condition holds, implying that strong duality holds between problem (6) and its dual. To obtain the latter, we write the dual function

$$d(\boldsymbol{\nu}) = \begin{cases} \min_{\boldsymbol{u}\in\mathbb{R}^\ell} \boldsymbol{b}^{\mathrm{T}}\boldsymbol{u} + \frac{\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}\|^2}{2} - \boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{u} & \text{if } \boldsymbol{\nu} \geq \boldsymbol{0} \\ -\infty & \text{otherwise.} \end{cases}$$

For $\boldsymbol{\nu} \geq \boldsymbol{0}$, the optimal solution of the minimization in $\boldsymbol{u}$ satisfies

$$\boldsymbol{b} - \boldsymbol{\nu} + \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}^* = \boldsymbol{0} \quad \Rightarrow \quad d(\boldsymbol{\nu}) = -\frac{1}{2}\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}^*\|^2.$$

Letting $\boldsymbol{x} = -\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}^* \in \mathbb{R}^m$, we get [3]

$$d(\boldsymbol{\nu}) = -\frac{\|\boldsymbol{x}\|^2}{2}, \quad -\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} = \boldsymbol{\nu}.$$

[3] Recall that $\|\boldsymbol{x}\| = \|-\boldsymbol{x}\|$ for any vector $\boldsymbol{x}$.

The dual problem $\text{maximize}_{\boldsymbol{\nu} \geq \mathbf{0}} \, d(\boldsymbol{\nu})$ can be written as

$$\begin{array}{ll} \text{maximize}_{\boldsymbol{\nu} \geq \mathbf{0}} & -\frac{\|\boldsymbol{x}\|^2}{2} \\ \text{s.t.} & -\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} = \boldsymbol{\nu}, \end{array}$$

that can be rewritten as a problem over $\boldsymbol{x}$ and $\boldsymbol{\nu}$

$$\begin{array}{ll} \text{maximize}_{\boldsymbol{x} \in \mathbb{R}^m, \boldsymbol{\nu} \in \mathbb{R}^\ell} & -\frac{\|\boldsymbol{x}\|^2}{2} \\ \text{s.t.} & -\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} = \boldsymbol{\nu} \\ & \boldsymbol{\nu} \geq \mathbf{0}. \end{array}$$

Eliminating further the variable $\boldsymbol{\nu}$, we finally arrive at

$$\begin{array}{ll} \text{maximize}_{\boldsymbol{x} \in \mathbb{R}^m} & -\frac{\|\boldsymbol{x}\|^2}{2} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}. \end{array}$$

d)ii) The dual problem (7) can be reformulated as

$$\underset{\boldsymbol{x} \in \mathbb{R}^m}{\text{minimize}} \, \frac{1}{2}\|\boldsymbol{x}\|^2 \quad \text{s.t.} \quad \boldsymbol{x} \in \mathcal{C}.$$

Consider the equivalent reformulation

$$\underset{\boldsymbol{x} \in \mathbb{R}^m}{\text{minimize}} \, \|\boldsymbol{x}\| \quad \text{s.t.} \quad \boldsymbol{x} \in \mathcal{C}.$$

The optimal value of this problem corresponds to the definition of the distance between the zero vector and the set $\mathcal{C}$, i.e.

$$\min_{\boldsymbol{x} \in \mathbb{R}^m} \{\|\boldsymbol{x}\| \mid \boldsymbol{x} \in \mathcal{C}\} = \text{dist}(\mathbf{0}, \mathcal{C}).$$

As a result,

$$\min_{\boldsymbol{x} \in \mathbb{R}^m} \frac{1}{2}\|\boldsymbol{x}\|^2 \quad \text{s.t.} \quad \boldsymbol{x} \in \mathcal{C} = \frac{1}{2}\text{dist}(\mathbf{0}, \mathcal{C})^2.$$

and the optimal value of problem (7) is given by $-\frac{1}{2}\text{dist}(\mathbf{0}, \mathcal{C})^2$.

d)iii) Since strong duality holds, the optimal value of problem (6) is $-\frac{1}{2}\text{dist}(\mathbf{0}, \mathcal{C})^2$, as is that of problem (4). Plugging this result into inequalities (4) and (3), we obtain

$$\ln\left(\mathbb{P}\left(\boldsymbol{y} \in \mathcal{C}\right)\right) \leq -\frac{1}{2}\text{dist}(\mathbf{0}, \mathcal{C})^2$$

and

$$\mathbb{P}\left(\boldsymbol{y} \in \mathcal{C}\right) \leq \exp\left[-\frac{1}{2}\text{dist}(\mathbf{0}, \mathcal{C})^2\right],$$

respectively.

NB: This is yet another concentration inequality.