DSO Screening Exam: 2017 In-Class Exam

Gregory Faletto

Exercise 1 (**Probability/Analysis**). (a)

$$\dot{Z}(t) = \frac{\partial}{\partial t}\sum_{i=1}^{k}\gamma_i(t)X_i = \sum_{i=1}^{k}\frac{\partial\gamma_i(t)}{\partial t}X_i$$

$$\implies \mathbb{E}\left(\dot{Z}(t)\right) = \sum_{i=1}^{k}\mathbb{E}\left(\frac{\partial\gamma_i(t)}{\partial t}X_i\right) = 0$$

$$\implies \text{Cov}(Z(t),\dot{Z}(t)) = \mathbb{E}\big[(Z(t)-\mathbb{E}[Z(t)])(\dot{Z}(t)-\mathbb{E}[\dot{Z}(t)])\big] = \mathbb{E}\big[Z(t)\dot{Z}(t)\big]$$

$$= \mathbb{E}\left[\left(\sum_{i=1}^{k}\gamma_i(t)X_i\right)\left(\sum_{i=1}^{k}\frac{\partial\gamma_i(t)}{\partial t}X_i\right)\right] = \sum_{i=1}^{k}\mathbb{E}\left(\gamma_i(t)\frac{\partial\gamma_i(t)}{\partial t}X_i^2\right) + 0 = \sum_{i=1}^{k}\gamma_i(t)\frac{\partial\gamma_i(t)}{\partial t}$$

(b)

$$\ddot{Z}(t) = \sum_{i=1}^{k}\frac{\partial^2\gamma_i(t)}{\partial t^2}X_i$$

Exercise 2 (**Mathematical statistics; hypothesis test. so maybe don't worry about?**). (a)

(b)

Exercise 3 (**Probability; Wen says we should do this**). (a) Let $u_j = \exp(\theta_j)$. In the $n = 1$ case we have

$$I(m,\lambda) = \int_{\mathbb{R}}\frac{\exp(\lambda\theta_1 m_1)}{(1+\exp(\theta_1))^{\lambda}}d\theta_1 = \int_{\mathbb{R}}\frac{\exp(\theta_1)^{\lambda m_1}}{(1+\exp(\theta_1))^{\lambda}}d\theta_1$$

Let $u_1 = 1+\exp(\theta_1) \implies \partial u_1 = \exp(\theta_1)\partial\theta_1$.

$$= \int_{\mathbb{R}}\frac{(u_1-1)^{\lambda m_1 - 1}}{u_1^{\lambda}}du_1$$

$$= \ldots = \frac{\Gamma(\lambda(1-m_1))\Gamma(\lambda m_1)}{\Gamma(\lambda)} = \frac{\Gamma(\lambda m_0)\Gamma(\lambda m_1)}{\Gamma(\lambda)}$$

(b)

Exercise 4 (**Convex Optimization**). **Like theorem 2 in convex optimization lecture notes 3; probably don't need to worry about since not covered in Boyd. Don't prioritize.**

Exercise 5 (**High-dimensional statistics**). (a) If $p > n$, then even if $\boldsymbol{X}$ is full rank, it has a nullspace with nonzero entries. That is, the columns of $\boldsymbol{X}$ must be linearly dependent, so there are infinitely many non-zero vectors $\boldsymbol{v}$ such that $\boldsymbol{X}\boldsymbol{v} = \boldsymbol{0}$). Therefore for any solution $\hat{\boldsymbol{\beta}}$ satisfying

$$\hat{\boldsymbol{\beta}} \in \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2^2$$

and any vector $\boldsymbol{v}$ in the nullspace of $\boldsymbol{X}$, $\hat{\boldsymbol{\beta}}+\boldsymbol{v}$ is also a solution since

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2^2 = \|\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}\|_2^2 = \|\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{0}\|_2^2 = \|\boldsymbol{y}-\boldsymbol{X}(\hat{\boldsymbol{\beta}}+\boldsymbol{v})\|_2^2$$

$$\iff \boldsymbol{\beta}+\boldsymbol{v} \in \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2^2.$$

(b) If we assume that $\|\boldsymbol{\beta}_0\|_0 = s \leq n$, then it may be that there is only one $\boldsymbol{v}$ in the nullspace of $\boldsymbol{X}$ such that $\|\hat{\boldsymbol{\beta}}\|_0 = s$. Intuitively, this makes sense because as long as the features with zero coefficients in $\boldsymbol{\beta}_0$ are sufficiently uncorrelated with the features with nonzero coefficients and the true coefficients are not too small, it is very unlikely that it is possible to construct another s-sparse vector with equally low empirical risk by replacing one of the true features with only one of the other features (or some combination thereof). (We would also hope that rank($\boldsymbol{X} > s$. Otherwise, it is either the case that features in the true model are linearly dependent, in which case the true model is not identifiable, or some of the noise features are linearly dependent with some of the true features, which could also preclude identifiability.)

(c) **Solution in Section 1.9.1 of Linear Regression notes.**

(d) **Solution in Section 1.9.1 of Linear Regression notes.**