

Seminários de Doutorado 2019/2 – Estatística em alta dimensão

Hugo Carvalho & Dani Gamerman

1 Proposta do seminário

Com o avanço na facilidade de coleta de dados através de mecanismos físicos cada vez mais sofisticados (e.g., máquinas de ressonância magnética, facilidades em sequenciamento genético, grandes telescópios, aceleradores de partículas), nos deparamos com um novo paradigma na Estatística. No cenário clássico onde temos $n > p$ (i.e., mais amostras do que parâmetros a serem estimados), tarefas como estimação de matrizes de covariância, regressão linear e PCA são garantidas de ter um bom desempenho através dos teoremas limite usuais da literatura de Teoria de Probabilidades. Porém, no cenário onde $p > n$, os mesmos métodos e garantias de convergência no caso clássico não são mais aplicáveis. Em particular, não podemos mais recorrer a resultados assintóticos.

O objetivo desse Seminário é dar uma breve introdução ao tópico de *Estatística em alta dimensão*, precisamente o estudo de técnicas de estimação e garantias de convergência no cenário $p > n$. A principal referência bibliográfica será o texto recém-lançado de Martin J. Wainwright – *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Ao longo dos 15 encontros previstos no semestre de 2019/2, visamos estudar, majoritariamente, os seguintes capítulos da referência:

- Capítulo 1: Introdução
- Capítulo 2: Desigualdades de concentração básicas
- Capítulo 6: Matrizes aleatórias e estimação de matrizes de covariância
- Capítulo 7: Modelos lineares esparsos em alta dimensão
- Capítulo 8: Análise de componentes principais em alta dimensão
- Capítulo 11: Modelos gráficos para dados em alta dimensão

Também poderão ser acrescentados outros materiais sobre o tema, trazido pelos participantes. Uma possível sub-divisão dos capítulos acima para as 15 apresentações é mostrada abaixo:

- 1) Capítulo 1
- 2) Capítulo 2: Seções 2.1 (desigualdades clássicas) e 2.2 (métodos baseados em martingais)
- 3) Capítulo 2: Seções 2.3 (funções Lipschitz de variáveis Gaussianas) e apêndices A e B (versões equivalentes de variáveis aleatórias sub-Gaussianas e sub-exponenciais, respectivamente)
- 4) Capítulo 6: Seções 6.1 (preliminares), 6.2 (matrizes de Wishart e seu comportamento)
- 5) Capítulo 6: Seções 6.3 (Matrizes de covariâncias de coleções sub-Gaussianas) e 6.4 (desigualdades para matrizes gerais)
- 6) Capítulo 6: Seção 6.5 (desigualdades para matrizes de covariância com alguma estrutura)
- 7) Capítulo 7: Seções 7.1 (formulação do problema e suas aplicações) e 7.2 (recuperação no caso sem ruído)

- 8) Capítulo 7: Seções 7.3 (estimação no cenário com ruído) e 7.4 (cotas no erro de predição)
- 9) Capítulo 7: Seção 7.5 (seleção de variáveis)
- 10) Capítulo 8: Seções 8.1 (Componentes principais e redução de dimensão) e 8.2 (desigualdades para autovetores)
- 11) Capítulo 8: Seção 8.3 (PCA esparsa)
- 12) Capítulo 11: Seções 11.1 (preliminares) e 11.2 (Estimação de modelos gráficos Gaussianos)
- 13) Capítulo 11: Seções 11.3 (modelos gráficos na forma exponencial) e 11.4 (modelos gráficos com variáveis corrompidas ou escondidas)
- 14)
- 15)

As duas apresentações livres destinam-se a apresentações complementares com materiais sugeridos pelos participantes, ou para um possível remanejamento do tempo dedicado a alguma parte especial do livro.