Chapter 1: Linear Models

1. Def 1.3: Linear model
2. Estimation of parameter b
3. $\hat{b}$ is unbiased estimator of b
4. Estimation of σ^2
5. Lemma 1.4: $\hat{\sigma^2}$ is unbiased estimator of σ^2
6. Theorem 1.5: $\hat{b}$ and $\hat{\sigma^2}$ are unbiased and uncorrelated, $\hat{b}$ is the best linear estimator of b and $\hat{\sigma^2}$ is the best quadratic estimator of σ^2
7. Lemma 1.6: some equalities for $y \sim \mathcal{N}_n(\mu, \Sigma)$, matricies $A, B \in \mathbb{R}^{n \times n}$
8. Remark: distribution properties of $\hat{b} - b$ and $\hat{\sigma^2}$.
9. Lemma 1.7: independence of linear and quadratic estimators.
10. Corollary 1.9: estimators $\hat{b}$ and $\hat{\sigma^2}$ are independent (when model is normally distributed).
11. Theorem 1.10: best unbiased estimator of $K^t b$
12. Theorem 1.11: definition of $Q \sim \mathcal{X}_r^2$
    1. Characteristic function
13. Corollary 1.12: law of $K^t \hat{b}$ and $\frac{n-k}{\sigma^2}\hat{\sigma^2}$
14. F-statistic
15. Theorem 1.16: F-test

Chapter 2: High Dimensional Linear Regression

1. $\hat{b}_{\text{naive}}$
2. LASSO estimator: $\hat{b}(\lambda)$
3. Remark 2.2: Duality, $\hat{b}_{\text{primal}}(r)$
4. Remark 2.3: Ridge regression,$\hat{b}_{\text{ridge}}(\lambda)$, $\hat{b}_{\text{ridge}_{primal}}(r)$
5. Soft and hard thresholds: $\hat{b}_j(\lambda), \hat{b}_{hard,j}(\lambda)$
6. Subdifferentials
7. Proposition 2.5: necessary and sufficient conditions for $\hat{b}(\lambda)$ to be LASSO
8. Lemma 2.6: Basic inequality
9. Defining set A
10. Proposition 2.7: defining λ_0 so that $\mathbb{P}[A] \approx 1$
11. Proposition 2.8: prediction error upper bound (**suboptimal**)
12. Inequalities (2.6) (rate of the prediction error, **suboptimal**) and (2.7) (deducing that the bound in Proposition 2.8 is suboptimal)
13. Defining vector b_S (2.8) (introducing zero/non-zero components vector of b)
14. Lemma 2.9: prediciton error upper bound in **reduced** model (using set $S_0 : \{j : b_{0j} \neq 0\}$)
15. Def. 2.10: **restricted** eigenvalue condition
16. Theorem 2.11: prediction error upper bound in **reduced** and **restricted** model
17. Corollary 2.12: rate of the prediction error in **reduced** and **restricted** model, **optimal**

M-Fold Crossvalidation: Data-driven Choise of λ

1. M-Fold crossvalidation algorithm
2. Def. 2.11: Adaptive LASSO: $\hat{b}_{\text{adapt}}(\lambda)$
3. LASSO vs Adaptive LASSO
    1. $\hat{S}_{\text{adapt}(\lambda)}, \mathbb{P}[\hat{S}_{\text{adapt}(\lambda)} = S_0] \to 1$ as $n \to \infty$, where S_0 - true number of non-zero elements.
    2. Typical LASSO estimate overestimates the number of non-zero components.
4. $\lambda_{\max} : \forall \lambda \geq \lambda_{\max}, \hat{b}(\lambda) = 0$

Chapter 3: Covariance, Correlation and PCA

1. Joint density in high dimensions
2. General definitions of MLE for μ and Σ
3. Theorem (3.1)
    1. MLE for μ
    2. MLE for Σ
4. Independence of $\hat{\mu}_{\text{ML}}$ and $\hat{\Sigma}_{\text{ML}}$
    1. Proposition 3.2: independence of linear transformations of X
    2. Proposition 3.3: independence of $\hat{\mu}_{\text{ML}}$ and $\hat{\Sigma}_{\text{ML}}$
5. Estimation of correlation coefficient
    1. Correlation definition
    2. When $corr(X, Y) = 1$
    3. Proposition 3.4: MLE for correlation
    4. MLE and bijective function
    5. CLT; Types of convergence; Slutsky's Lemma; Delta method
    6. Theorem (3.5): CLT for: $\hat{\mu}_n$; $\hat{\Sigma}_n$; $\hat{p}_{ij}$
6. PCA
    1. Def. 3.6: PCA
    2. Theorem 3.7: ML-estimators of λ and β
        1. Mapping $\Sigma \to (\lambda_1, \ldots, \lambda_k, \beta_1, \ldots, \beta_k)$ is bijective
    3. Theorem 3.8

Chapter 4: Estimation of Large Covariance Matricies

Sparse covariance matricies

1. Introduction, notations
2. Remark: what can go wrong in high dimension

3. Theorem 4.1: error between sparse hard threshold estimator of Σ and Σ
    1. Stochastic order
    2. Remark after proof

Covariance matricies in band-form

1. Introduction, notations
2. Theorem 4.2