SubGaussian RVs

1. SubGaussian RVs
2. Orlicz norm $|| \cdot ||_{\psi_2}$
3. Hoeffding's Inequality (Version 1 and Version 2)

SubExponential RVs

1. SubExponential RVs
2. Orlicz norm $|| \cdot ||_{\psi_1}$
3. $|| \cdot ||_{\psi_2}$ vs $|| \cdot ||_{\psi_1}$
4. Centering Lemma
5. Bernstein's inequality

PCA

1. PCA
2. Theoretical PCA problem
3. Empirical covariance matrix estimator
4. PCA strategy
5. PCA summary

Control of Eigenvalues and Eigenvectors

1. Covariance matrix
2. Operator norm
3. Weyl's Theorem - eigenvalues
4. Daki's - Kahane Theorem - eigenvectors
    1. Frobenious norm

Control of Maxima (SubGaussian)

1. $\mathbb{E}[\max X_i]$ and $\mathbb{P}[\max X_i > t]$ for X_i in infinite set
2. Maximum over a convex polytope (finite set)
    1. Convex polytope
    2. B_1^d is a convex polytope
    3. Lemma: maximum for an infinite set = maximum for a convex polytope (finite set)
    4. $\mathbb{E}[\max X_i]$ and $\mathbb{P}[\max X_i > t]$ for X_i in a convex polytope (finite set)
3. Maximum over B_2^d of $\mathcal{R}^d$
    1. Relationship between L_2^d and L_1^d balls (sub-optimal upperbound)
    2. ε - net
    3. Lemma that B_2^d admits ε - net and cardinality of such ε - net
    4. Algorithm to build ε - net of B_2^d
    5. Maximum over

Operator norm control

1. Upperbound on the operator norm of $\hat{\Sigma} - \Sigma$

Kernel Density Estimation

1. Density
    1. Radon-Nikodym Theorem (existence of density)
2. Density and Likelihood
    1. Density vs Likelihood
    2. Maximum likelihood estimator
3. Kernel Density Estimator
    1. Convolution product
    2. Convolution properties
    3. Approximate identity
    4. Approximate identity in C^p
    5. KDE
    6. KDE Bandwidth
    7. Impact of the convolution

Statistical perfomance of KDE

1. Risk of KDE: $MSE_h(x_0)$ and $MSE_h = MISE_h$
2. Bias and variance of KDE
3. Upperbounding the variance
4. Upperbounding the bias
    1. Taylor expansion of f at x_0
    2. Kernel of order K
    3. Hölder class
5. Upperbounding $MSE_h(x_0)$
    1. Finding h_{opt} that achieves smallest upperbound
6. Upperbounding MSE_h with h_{opt}

Data-driven calibration of h

1. Estimating $MSE(h)$ with cross-validation
2. h_{oracle} - h that achievest the smallest $MSE(h)$
3. Leave-one-out cross-validation
4. Estimating $L(h)$ with leave-one-out estimator and $\hat{h}_{L1O}$
5. Empirical risk for h.

PCO strategy

1. $MSE(h) = b_h + V_h$
    1. Estimating b_h
    2. Estimating V_h
2. PCO strategy: $\hat{h}_{\text{PCO}}$
3. PCO oracle inequality

Empirical risk minimization (ERM)

1. Statistical learning problem
    1. Supervised setting
        1. Binary classification

2. Regression
    1. Linear regression
    2. Constrained linear regression
    3. Single-index model
    4. Multi-index model
    5. Deep NNs
2. Unsupervised setting
    1. Density estimation
        1. Log-likelihood & Kullback-Leibler divergence
2. Empirical risk minimization
    1. Risk (expectation of loss)
    2. Excess risk (risk of a candidate function minus risk of the best possible function)
    3. Empirical risk (instead of expectation => $\frac{1}{n} \sum_{i=1}^{n}$ of loss for all X_i's)

Statistical perfomance of ERM

1. Trying to upperbound excess risk of ERM
2. Komolgorov-Smirnov statistic
3. Glivenko-Cantelli class
4. Upperbounding excess risk of ERM for finite cardinality of candidate functions f: $|\mathcal{F}| < +\infty$
5. Upperbounding excess risk of ERM for infinite cardinality of candidate functions f: $|\mathcal{F}| = +\infty$ => Rademacher complexities

Rademacher complexities

1. Rademacher complexity
2. Empirical Rademacher complexity of $\mathcal{F}(X_1^n)$
3. Rademacher complexity of the class $\mathcal{F}$
4. Properties of Rademacher complexity
5. Basic properties of Rademacher complexity
6. Interpretation
7. Bounded-difference inequality (BDI)
8. Concentration property of Rademacher complexity
9. Upperbounding a class of bounded functions using Rademacher complexity (tight result)

Binary classification

1. Notations

2. Lemma for $Z_i = (X_i, Y_i)$, set of binary classifiers H and set of losses G and empirical Rademacher complexity
3. Corollary: upperbounding the difference between the risk and empirical risk over a set of binary classifiers

Upperbounding Rademacher complexity

1. Massart theorem (upperbound on the rademacher complexity of a set)

VC-classes and VC-dimension

1. VC-theorem
2. VC-dimension
    1. Shatter coefficient of order n
    2. VC-dimension
    3. Sauer's theorem
        1. Corollary
3. Connection between VC-classes and sets (proving that H (set of binary classifiers) is a VC-class)
    1. Corollary: VC-sets (upperbounding the difference between the risk of a classifier h and its empirical risk)

SVM

1. Reproducing kernels and SVM
    1. Positive semi-definite (PSD) and PD kernels
    2. Why use PSD kernels?
    3. Properties (propositions + corollary) for PSD kernels and PSD kernel types (polynomial, exponential, normalized, Gaussian (RBF))
2. Reproducing kernel Hilbert space (RKHS)
    1. Reproducing kernel
    2. Hilbert space properties (+ Cauchy-Schwarz for Hilber space)
    3. Kernel trick
    4. Bounded kernel
    5. All functions of Hilbert space are bounded