CSC 665: Final exam topics

Chicheng Zhang

December 8, 2019

Topics in the first half (20%):

1. Concentration Inequalities
   (a) Chernoff bound
   (b) Hoeffding's Inequality
   (c) McDiarmid's Lemma
   (d) Application to evaluation / validation of classifiers
2. PAC learning for iid binary classification
   (a) Setup of PAC learning model
   (b) Definition of sample complexity, PAC learnability, Agnostic PAC learnability
   (c) Sample complexity of ERM in realizable and agnostic settings for finite classes
3. VC theory
   (a) Definition of VC dimension
   (b) Examples of VC dimension for simple function classes (e.g. threshold, linear classes)
   (c) Sauer's Lemma
4. Rademacher complexity and uniform convergence
   (a) Definition of Rademacher complexity
   (b) Rademacher complexity for 0-1 loss induced by VC function classes
   (c) Uniform convergence of empirical error to generalization error for VC classes
5. Sample complexity lower bound for PAC learning
   (a) Sample complexity lower bound for hypothesis class of VC dimension d
   (b) Fundamental Theorem of Statistical Learning

Topics in the second half (80%):

1. Support Vector Machines
   (a) The SVM formulation
   (b) KKT condition; Interpretation of support vectors
   (c) Margin-based generalization bounds (abbrev. margin bounds): theorem statement;
   (d) Why SVM works well in practice: if data is linearly separable by a margin, derive a generalization error bound for SVM solution

2. Boosting
   (a) The AdaBoost algorithm
   (b) Margin bounds; comparison between ℓ_1 / ℓ_∞ margin bounds and ℓ_2 / ℓ_2 margin bounds - in what settings is one better than the other?
3. Model selection
   (a) Error decomposition in ML: generalization error, optimization error, bias, and methods to detect and control each
   (b) Model selection: cross valiation and structural risk minimization
4. Online classification
   (a) Online learning: problem formulations
   (b) Realizable online classification: the mistake bound model
   (c) Consistency, Halving and their analysis
   (d) Littlestone's dimension: examples for simple function classes (e.g. thresholds or finite classes); comparison with VC dimension
   (e) Standard optimal algorithm and its performance guarantees
5. Prediction with expert advice (PEA)
   (a) Using PEA for online classification
   (b) Definition of regret
   (c) Cover's impossiblity result: no deterministic algorithm can get sublinear regret
   (d) Using decision-theoretic online learning (DTOL) for PEA
   (e) The Hedge algorithm and its $O(\sqrt{T \ln N})$ regret analysis; why is it better than follow the leader (FTL)?
6. Online to batch conversion
   (a) Statement: running bounded-regret algorithms for iid sequences, Regret guarantee implies excess loss guarantee
   (b) Two methods for conversion: (1) returning a predictor uniformly at random; (2) when the loss is convex, return the averaged predictor over history.
7. Convexity
   (a) Norms and their dual norms; examples (e.g. ℓ_p vs. ℓ_q for $\frac{1}{p} + \frac{1}{q} = 1$, and $\|\cdot\|_A$ vs. $\|\cdot\|_{A^{-1}}$.)
   (b) Convex sets and convex functions
   (c) Subgradients (for functions that are not-necessarilty differentialble, such as $f(w) = |w|$ or $(1 - w)_+$); first-order-approximation using subgradients must lie below the original convex function
   (d) Bregman divergence; its nonnegativity and asymmetry
   (e) Lipschitzness; equivalence to bounded subgradient
   (f) Strong convexity and smoothness (e.g. checking strong convexity and smoothness parameter for function $\frac{1}{2}\|x\|_2^2$); relationship to Bregman divergence
   (g) Legendre-Fenchel duality (and its geometrical explanation); strong convexity implies smoothness of the dual

8. Online convex optimization (OCO)
   (a) Setup; regret against a point vs. regret against a convex set
   (b) Online linear optimization (OLO)
   (c) Follow the regularized leader (FTRL) for OLO: algorithm and performance guarantees $O(\sqrt{T})$ regret; examples of FTRL with different regularizers and guarantees
   (d) Solving OCO by solving OLO
   (e) Online classification by OLO: Perceptron and Winnow; their performance guarantees in realizable / nonrealizable settings
   (f) Adaptive regularization for OLO: algorithm and performance guarantees
   (g) Online strongly-convex optimization: $O(\ln T)$ regret; implication for SVM optimization