Errata Relevant to Machine Learning Theory Course for ‘Understanding Machine Learning: From Theory to Algorithms”

Tim van Erven

November 10, 2021

Chapter 2

- Ex. 2.3, hint: Rx cannot be taken to have exactly probability epsilon if there are point-masses.

Chapter 6

- Lemmas A.5 and 6.10 only show the polynomial bound on the growth function for $m > d + 1$, but Theorem 6.11 uses that bound for $m > d$. This can be fixed by replacing the long proof of Lemma A.5 with the much simpler proof from the book of Anthony and Bartlett that was shown in the lecture, which holds for $m > d$.
- Theorems 6.7, 6.8 and Chapter 28 require VC-dimension $d \geq 1$, and additional care is also required for existence of a truly universal constant C because ϵ and δ may be arbitrarily close to 1, so then we cannot hide any additive constants. E.g. there may not exist any C' such that $\ln(4/\delta) <= C' \ln(1/\delta)$. This is no problem if we replace $\ln(1/\delta)$ in all theorem statements by $\ln(e/\delta)$.
- Theorem 6.11 is missing a “for all $h \in \mathcal{H}$” statement.
- The proof of Theorem 6.11 uses Hoeffding's inequality for independent random variables, but it is only proved for i.i.d. random variables in the appendix. This is not a real problem, because the proof for Hoeffding's inequality also goes through in the same way for independent random variables.
- Ex. 6.11: log should probably be $\log_2$

Chapter 7

- Definition 7.1 and below: The definition of agnostic PAC-learning is missing the requirement that the learner is proper (i.e. $h_S \in \mathcal{H}$). It is unclear from the definition of non-uniform learnability whether there is such a requirement, but we can assume that there is.

- Theorem 7.2: Second half of proof ("For the other direction...") is incorrect: it is argued that $\mathcal{H}_n$ is PAC-learnable and hence must have finite VC-dimension by the fundamental theorem of PAC-learning. But the argument that $\mathcal{H}_n$ is PAC-learnable is incomplete for two reasons:

  1. PAC-learnability is a requirement for all ϵ, and δ, but here there is only an argument for $\delta = 1/7$ and $\epsilon = 1/8$.
  2. PAC-learnability requires a proper learner, which is not guaranteed for the learning algorithm A.

  The fix for both these issues is that, instead of referring to the fundamental theorem, the proof should refer to Corollary 6.4, which solves both problems. In addition, to get the desired implication from Corollary 6.4, it is necessary to take smaller values $\delta < 1/7$ and $\epsilon < 1/8$ (with strict inequality). Any choice will do; for instance, $\delta = 1/14$ and $\epsilon = 1/16$.

Chapter 26

- Lemma 26.6: the inequality actually holds with equality.
- Lemma 26.9: $\phi_m(y_m)$ should be $\phi_m(a_m)$.

Chapter 28

- Section 28.1: the bound in (28.1) can easily be strengthened to

  $$m_{\mathcal{H}}(\epsilon, \delta) \leq C \frac{d \ln(e/\epsilon) + \ln(1/\delta)}{\epsilon^2},$$

  i.e. without the d inside the logarithm. This can be done by replacing the last step in the analysis, where the book appeals to Lemma A.2, by the argument from homework 5, which upper bounds the concave logarithm $\ln(m)$ by its tangent at a well-chosen point m_0: $\ln(m) \leq \ln(m_0) + \frac{m - m_0}{m_0}$.