Notations for CS 578: Statistical Machine Learning (keep updating)

Ruizhe Zhang

General notations. We use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. We use the standard asymptotic notations (e.g., big-O, big-Ω, etc.)[1]. We use $\widetilde{O}$ to hide polylogarithmic factors in problem parameters. We use $\simeq$, $\lesssim$, and $\gtrsim$ to hide universal constants, e.g., $x \lesssim y$ means there is a universal constant C such that $x \leq Cy$. We use $\mathbf{1}_{\mathcal{E}}$ or $\mathbf{1}\{\mathcal{E}\}$ to denote the indicator function for the Boolean predicate $\mathcal{E}$; that is, $\mathbf{1}_{\mathcal{E}} = 1$ when $\mathcal{E}$ is true, and $\mathbf{1}_{\mathcal{E}} = 0$ when $\mathcal{E}$ is false.

Probability. We use $\mathbb{E}$ to denote the expectations of random variables with respect to all randomness used to define the variable unless otherwise specified. We use $\mathrm{Var}[X] := \mathbb{E}[X^2] - \mathbb{E}[X]^2$ to denote the variance of X. For a probability distribution $\mathcal{D}$, we write $x \sim \mathcal{D}$ to denote a sample from this distribution. We use $\mathcal{N}(\mu, \sigma^2)$ to note the 1-d Gaussian distribution with expectation μ and variance σ^2.

[1] https://en.wikipedia.org/wiki/Big_O_notation