CheatSheet Basics Machine Learning

Chu Duc Thang

April 2021

1 Chapter 1

Introduction to the class. Only 1 page, no important information here

2 Chapter 2: Probability

1. Sample space/outcome space vs event space:
   - Sample space: Ω
   - Event space: Subset of sample space, ex: powerset (discrete), Borel Field (continuous)
2. Discrete vs Continuous RV
   - Discrete: {}, N, words
   - Continuous: [], R, R^k
3. Probability mass function (pmf) vs probability density function (pdf)
   - Pmf: $\Omega \to [0, 1]$
   - Pdf: $\Omega \to [0, \text{inf})$, no singleton event, can be > 1
4. Special Distribution
   - Discrete: Uniform (n - #outcomes), Poisson (α - histogram/likely), Bernoulli (p - success)
   - Continuous: Gamma (α, β), Uniform (a, b), Normal(μ, σ), Exponential (α)
5. Marginal vs Conditional Distribution
   - Marginal: p(x) = $\sum_{y \in Y} p(x, y)$
   - Conditional: p(x‖y) = $\frac{p(y\|x)p(x)}{p(y)}$ or p(x,y,z) = p(x‖y,z)p(y‖z)p(z)
6. Expected value vs Conditional Expected value vs Variance

- E = $\sum_{x \in X} xp(x)$
- E[X‖Y] = $\sum_{x \in X} xp(x\|y)$
- Var = E[$(X - E[X])^2$] or $E[X^2] - E[X]^2$
- Properties of E: E[c] = c, E[cX] = cE[X], E[X + Y] = E[X] + E[Y], E[XY] = E[X]E[Y] (independence), E[E[Y‖X]] = E[Y]
- Properties of Var: Var[c] = 0, Var[cX] = c^2Var[X], Var[X + Y] = Var[X] + Var[Y] + 2Cov(X,Y)

7. Covariance vs Correlation
   - Cov = E[XY] - E[X]E[Y]
   - Corr = $\frac{Cov(x,y}{\sqrt{Var(x)}\sqrt{Var(y)}}$
   - Note: $-1 \leq Corr \leq 1$, but Cov is unbounded
8. Independence vs Conditional Independence
   - P(X,Y) = P(X)P(Y)
   - P(X,Y‖Z) = P(X‖Z)P(Y‖Z)

3 Chapter 3: Estimator

4 Chapter 4: Optimization

5 Chapter 5: MAP/MLE/Bayesian

6 Chapter 6: Optimal predictor

7 Chapter 7: Linear/Polynomial Regression

8 Chapter 8: Generalization Error

9 Chapter 9: Regularization

10 Chapter 10: Classification