Formula Sheet

Chapter 10: Generalization and Bias-Variance

- Bias-Variance Decomposition:

$$\mathbb{E}[(f_{\mathcal{D}}(X) - f^*(X))^2] = \mathbb{E}_X \left[(\mathbb{E}_{\mathcal{D}}[f_{\mathcal{D}}(X)] - f^*(X))^2 + \mathrm{Var}_{\mathcal{D}}[f_{\mathcal{D}}(X)]\right] \tag{1}$$

- Double Descent Phenomenon: Occurs in overparameterized regimes where test error decreases as model capacity increases beyond interpolation threshold.

Chapter 11: Mixture Models

- Mixture Model Density:

$$p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{k=1}^{m} w_k p(\mathbf{x}|\boldsymbol{\theta}_k) \tag{2}$$

- EM Algorithm:

$$\text{E-step: } p_t[i,k] = \frac{w_k^{(t)} p(\mathbf{x}_i|\boldsymbol{\theta}_k^{(t)})}{\sum_{j=1}^{m} w_j^{(t)} p(\mathbf{x}_i|\boldsymbol{\theta}_j^{(t)})} \tag{3}$$

$$\text{M-step: } w_k^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} p_t[i,k], \quad \boldsymbol{\theta}_k^{(t+1)} = \mathrm{argmax}_{\boldsymbol{\theta}_k} \sum_{i=1}^{n} p_t[i,k] \ln p(\mathbf{x}_i|\boldsymbol{\theta}_k) \tag{4}$$

- Gaussian Mixture Updates:

$$\boldsymbol{\mu}_k = \frac{\sum_{i=1}^{n} p_t[i,k]\mathbf{x}_i}{\sum_{i=1}^{n} p_t[i,k]}, \quad \boldsymbol{\Sigma}_k = \frac{\sum_{i=1}^{n} p_t[i,k](\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top}{\sum_{i=1}^{n} p_t[i,k]} \tag{5}$$

Chapter 12: Variational Autoencoders (VAEs)

- Evidence Lower Bound (ELBO):

$$\mathrm{ELBO} = \mathbb{E}_{\mathbf{h}\sim q}[\ln p(\mathbf{x}|\mathbf{h},\mathbf{W})] - D_{\mathrm{KL}}(q(\mathbf{h}|\mathbf{x})\|p(\mathbf{h})) \tag{6}$$

- Reparameterization Trick:

$$\mathbf{h} = \boldsymbol{\mu}(\mathbf{x}) + \boldsymbol{\sigma}(\mathbf{x}) \odot \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0},\mathbf{I}) \tag{7}$$

- KL Divergence for Gaussian Encoder:

$$D_{\mathrm{KL}}(q(\mathbf{h}|\mathbf{x})\|\mathcal{N}(\mathbf{0},\mathbf{I})) = \frac{1}{2}\sum_{j=1}^{d}\left(\sigma_j^2 + \mu_j^2 - 1 - \ln \sigma_j^2\right) \tag{8}$$

- Conditional VAE ELBO:

$$\mathrm{ELBO}_{\mathrm{cond}} = \mathbb{E}_{\mathbf{h}\sim q}[\ln p(\mathbf{x}|\mathbf{h},\mathbf{y},\mathbf{W})] - D_{\mathrm{KL}}(q(\mathbf{h}|\mathbf{x},\mathbf{y})\|\mathcal{N}(\mathbf{0},\mathbf{I})) \tag{9}$$

Chapter 13: Evaluating Generative Models

- Test Log-Likelihood:

$$\text{Perf}(\theta) = \frac{1}{m} \sum_{\mathbf{x} \in \mathcal{D}_{\text{test}}} \ln p(\mathbf{x}|\theta) \tag{10}$$

- Importance Sampling for VAE Likelihood:

$$p(\mathbf{x}|\mathbf{W}) \approx \frac{1}{m} \sum_{k=1}^{m} \frac{p(\mathbf{h}_k)}{q(\mathbf{h}_k|\mathbf{x})} p(\mathbf{x}|\mathbf{h}_k, \mathbf{W}), \quad \mathbf{h}_k \sim q(\mathbf{h}|\mathbf{x}) \tag{11}$$

Chapter 14: Missing Data

- Matrix Completion Objective:

$$\min_{\mathbf{H},\mathbf{D}} \sum_{i=1}^{n} \sum_{j \in \mathcal{A}_i} (x_{ij} - \mathbf{h}_i \mathbf{D}_{:,j})^2 \tag{12}$$

- Autoencoder Imputation Loss:

$$\min_{\mathbf{W}, \mathbf{x}_{\mathcal{M}_i}} \sum_{i=1}^{n} \|\mathbf{x}_i - f_{\mathbf{W}}(\mathbf{x}_i)\|_2^2 \tag{13}$$

Chapter 15: Bayesian Methods

- Bayesian Linear Regression Posterior (Multivariate):

$$\boldsymbol{\mu}_n = \boldsymbol{\Lambda}_n^{-1} \mathbf{X}^\top \mathbf{y}, \quad \boldsymbol{\Sigma}_n = \sigma_y^2 \boldsymbol{\Lambda}_n^{-1}, \quad \boldsymbol{\Lambda}_n = \mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I} \tag{14}$$

- Gaussian Process Regression:

$$\mathbb{E}[f(\mathbf{x})] = \mathbf{k}(\mathbf{x}, \mathbf{X})(\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{y} \tag{15}$$

$$\text{Var}[f(\mathbf{x})] = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}(\mathbf{x}, \mathbf{X})(\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{k}(\mathbf{x}, \mathbf{X})^\top \tag{16}$$

- Normal-Inverse-Gamma (NIG) Posterior Updates:

$$\boldsymbol{\Lambda}_n = \mathbf{X}^\top \mathbf{X} + \boldsymbol{\Lambda}_0, \quad \boldsymbol{\mu}_n = \boldsymbol{\Lambda}_n^{-1}(\mathbf{X}^\top \mathbf{y} + \boldsymbol{\Lambda}_0 \boldsymbol{\mu}_0) \tag{17}$$

$$a_n = a_0 + \frac{n}{2}, \quad b_n = b_0 + \frac{1}{2}\left(\mathbf{y}^\top \mathbf{y} + \boldsymbol{\mu}_0^\top \boldsymbol{\Lambda}_0 \boldsymbol{\mu}_0 - \boldsymbol{\mu}_n^\top \boldsymbol{\Lambda}_n \boldsymbol{\mu}_n\right) \tag{18}$$