Machine Learning – Key Concepts & Formula Sheet

For BITS AIML CZG565 (Exam-Oriented Summary)

Module 1: Introduction to Machine Learning

- Learning task defined by ■Task (T), Performance (P), Experience (E)■
- Supervised, Unsupervised, Semi-supervised, Reinforcement Learning
- Classification → categorical output, Regression → continuous output
- Generative models learn $P(x,y)$; Discriminative models learn $P(y|x)$

Module 2: ML Workflow & Data Preprocessing

- Workflow: Problem → Data → Cleaning → Features → Model → Train → Evaluate
- Data issues: Missing values, noise, outliers, duplicates
- $IQR = Q3 – Q1$; Outliers if $x < Q1 – 1.5\times IQR$ or $x > Q3 + 1.5\times IQR$
- 3-Sigma Rule: $\mu \pm 3\sigma$
- Min–Max: $(x – xmin)/(xmax – xmin)$
- Z-score: $(x – \mu)/\sigma$

Module 3: Linear Regression

- Model: $\blacksquare = w0 + w1x1 + \ldots + wnxn$
- MSE Cost: $J = (1/n)\ \Sigma(\blacksquare – y)^2$
- Gradient Descent: $wj = wj – \alpha(\partial J/\partial wj)$
- Normal Equation: $w = (X\blacksquare X)\blacksquare^1 X\blacksquare y$
- $R^2 = 1 – SSres/SStot$

Regularization

- Ridge (L2): $J = MSE + \lambda\Sigma wj^2$

$\partial J/\partial wj = (1/n)\Sigma(\blacksquare – y)xj + (\lambda/n)wj$
- Lasso (L1): $J = MSE + \lambda\Sigma|wj|$

$\partial J/\partial wj = (1/n)\Sigma(\blacksquare – y)xj + (\lambda/n)sign(wj)$

Module 4: Logistic Regression

- Sigmoid: $\sigma(z) = 1 / (1 + e\blacksquare\blacksquare)$
- Model: $h(x) = \sigma(w\blacksquare x)$
- Cross-Entropy Cost:

$J = –(1/n)\Sigma[y\log(h) + (1–y)\log(1–h)]$
- Decision rule: $h(x) \geq 0.5$ → class 1
- $logit(p) = \log(p/(1–p)) = w\blacksquare x$
- Accuracy = (TP+TN)/Total
- Precision = TP/(TP+FP)
- Recall (TPR) = TP/(TP+FN)
- FPR = FP/(FP+TN)
- ROC curve plots TPR vs FPR; AUC=1 perfect, 0.5 random

Module 5: Decision Trees

- Entropy: $H(S) = –\Sigma pi \log\blacksquare\ pi$
- Information Gain: $Gain(S,A) = Entropy(S) – \Sigma(|Sv|/|S|)Entropy(Sv)$
- Gini Index: $1 – \Sigma pi^2$
- Gain Ratio = Gain / SplitInfo

- Stop when node is pure, no attributes left, or gain is small
- Overfitting control: pre■pruning, post■pruning, depth limits