Machine Learning Revision Summary

1. 1. Basics - Regression vs Classification

English	Chinese (中文)
Question: Give an example of a (1) regression and (2) classification problem.	问题: 举例说明 (1) 回归问题 和 (2) 分类问题。
Solution: Regression: Predicting the average global temperature in 2050. (Target is continuous/ordinal) Classification: Identifying the species of an Isopod from an image.	解答: 回归：预测2050年的全球平均气温。（目标是连续或有序的） 分类：根据图像识别等足虫的种类。

2. 1. Basics - Extrapolation vs Interpolation

English	Chinese (中文)
Question: Give examples of Extrapolation vs Interpolation.	问题: 举例说明外推 (Extrapolation) 与内插 (Interpolation) 的区别。
Solution: Extrapolation: Weather forecasting (predicting outside known range). Interpolation: Predicting a student's grade from their attendance, given the rest of the class's attendance and grades (predicting within known range).	解答: 外推：天气预报（在已知范围之外进行预测）。 内插：根据全班的出勤率和成绩，通过某学生的出勤率预测其成绩（在已知范围内进行预测）。

3. 1. Basics - Circular Analysis

English	Chinese (中文)
Question: Define Circular Analysis and give an example.	问题: 定义循环分析 (Circular Analysis) 并举例。
Solution: Definition: Selecting analysis details (e.g., parameters) using the data that is being used for testing. Example: Adjusting preprocessing parameters on fMRI data to get the 'best' result for the test data, leading to falsely significant results.	解答: 定义：使用用于测试的数据来选择分析细节（例如参数）。 示例：调整fMRI数据的预处理参数以获得测试数据的“最佳”结果，导致结果虚假显著。

4. 2. Entropy - Definition

English	Chinese (中文)
Question: Write down the equation that defines the entropy of a discrete random variable, X.	问题: 写出定义离散随机变量 X 的熵的公式。
Solution: $H[X] = E[-\log p(X)] = - \sum_{i=1}^{N_X} p(x_i) \log p(x_i)$	解答: $H[X] = E[-\log p(X)] = - \sum_{i=1}^{N_X} p(x_i) \log p(x_i)$

5. 2. Entropy - Conditional Entropy

English	Chinese (中文)
Question: Define conditional entropy in words.	问题: 用文字定义条件熵 (Conditional Entropy)。
Solution: The conditional entropy of X given Y is the expectation (over Y) of the entropy of X given the values of Y. It is the sum of H(X|Y=y) weighted by P(Y=y).	解答: 给定 Y 时 X 的条件熵是 X 给定 Y 值时的熵的（关于 Y 的）期望。 它是 H(X|Y=y) 加权 P(Y=y) 的和。

6. 2. Entropy - Information Gain

English	Chinese (中文)
Question: Define information gain.	问题: 定义信息增益 (Information Gain)。
Solution: It is the expected reduction in entropy given new information: $IG(X, Y) = H(X) - H(X\|Y)$.	解答: 它是给定新信息后熵的预期减少量：$IG(X, Y) = H(X) - H(X\|Y)$。

7. 3. Validation - Timeseries Split

English	Chinese (中文)
Question: Traffic speeds are collected every minute. A random cross-validation approach shows very good prediction from weather. (a) What is the problem? (b) How to address it?	问题: 每分钟收集交通速度。随机交叉验证显示现在的天气预测效果很好。(a) 有什么问题？ (b) 如何解决？
Solution: (a) Neighbouring minutes are highly correlated. Random splitting puts correlated samples in both train and test sets, leading to data leakage/over-optimistic results. (b) Use a time-series split (e.g., train on past, test on future) or split by day, ensuring independence.	解答: (a) 相邻分钟的数据高度相关。随机拆分会导致相关的样本同时出现在训练集和测试集中，导致数据泄漏/结果过于乐观。 (b) 使用时间序列拆分（例如，以前的数据训练，以后的数据测试）或按天拆分，确保独立性。

8. 4. Decision Trees - Purity

English	Chinese (中文)
Question: Explain the general idea of how we use 'purity' to build a decision tree.	问题: 解释如何在构建决策树时使用“纯度 (Purity)”这一概念。
Solution: We pick a feature to split the data such that the resulting child nodes are as 'pure' as possible (contain mostly one class or low variance). We maximize Information Gain (reduction in impurity).	解答: 我们选择一个特征来分割数据，使得生成的子节点尽可能“纯”（主要包含一个类别或方差较小）。我们需要最大化信息增益（不纯度的减少）。

9. 5. Ensemble - Bagging (Bootstrap Aggregation)

English	Chinese (中文)
Question: How to use bagging to compute a 95% confidence interval for a neural network prediction?	问题: 如何使用 Bagging 计算神经网络预测的 95% 置信区间？
Solution: 1. Resample training data with replacement multiple times (bootstrap samples). 2. Train a network on each sample. 3. Predict on the test point with all networks to get a distribution. 4. Take the 2.5th and 97.5th percentiles of the predictions.	解答: 1. 对训练数据进行多次有放回重采样（Bootstrap 样本）。 2. 在每个样本上训练一个网络。 3. 使用所有网络对测试点进行预测，得到一个分布。 4. 取预测值的第 2.5 和第 97.5 百分位。

10. 5. Ensemble - Random Forest

English	Chinese (中文)
Question: Explain the Random Forest algorithm steps.	问题: 解释随机森林 (Random Forest) 算法的步骤。
Solution: 1. Generate multiple bootstrap samples. 2. Build a decision tree for each sample. 3. At each split, consider only a random subset of features (subspace sampling). 4. Aggregate predictions (vote for classification, average for regression).	解答: 1. 生成多个 Bootstrap 样本。 2. 为每个样本构建一棵决策树。 3. 在每个分裂点，只考虑随机选择的一部分特征（子空间采样）。 4. 聚合预测结果（分类用投票，回归取平均）。

11. 6. Linear Regression - Cost Functions

English	Chinese (中文)
Question: Write the Sum Squared Error (SSE) in matrix notation.	问题: 用矩阵符号写出平方误差和 (SSE)。
Solution: $SSE = (y - Xw)^T(y - Xw)$	解答: $SSE = (y - Xw)^T(y - Xw)$

12. 6. Linear Regression - Regularization

English	Chinese (中文)
Question: What term is added for L2 regularization (Ridge)?	问题: L2 正则化 (Ridge) 增加了什么项？
Solution: Add $\lambda w^T w$ (sum of squared weights). Penalizes large weights.	解答: 增加 $\lambda w^T w$（权重的平方和）。惩罚较大的权重值。

13. 7. Gaussian Processes - Definition

English	Chinese (中文)
Question: Define Gaussian Process.	问题: 定义高斯过程 (Gaussian Process)。
Solution: A stochastic process where any finite collection of random variables has a multivariate normal distribution. Defined by a mean function and a kernel (covariance) function.	解答: 一种随机过程，其中任意有限个随机变量的集合都服从多元正态分布。由均值函数和核（协方差）函数定义。

14. 7. Gaussian Processes - Hyperparameters

English	Chinese (中文)
Question: How to handle uncertainty in lengthscale?	问题: 如何处理长度尺度 (Lengthscale) 的不确定性？
Solution: Use a Bayesian approach: Place a prior on the lengthscale, compute the marginal likelihood, and integrate it out (or approximate via sampling/MCMC) instead of just using a point estimate.	解答: 使用贝叶斯方法：为长度尺度设置先验，计算边际似然，并将其积分掉（或通过采样/MCMC近似），而不是仅使用点估计。

15. Worksheet - Logistic Regression

English	Chinese (中文)
Question: Derive the negative log-likelihood for Logistic Regression (Bernoulli).	问题: 推导逻辑回归（伯努利）的负对数似然。
Solution: Given $P(y\|x) = \pi^y (1-\pi)^{1-y}$. Log-likelihood $L = \sum [y_i \log \pi_i + (1-y_i) \log(1-\pi_i)]$. Negative Log-Likelihood is the negative sum of these terms (Cross Entropy).	解答: 给定 $P(y\|x) = \pi^y (1-\pi)^{1-y}$。 对数似然 $L = \sum [y_i \log \pi_i + (1-y_i) \log(1-\pi_i)]$。 负对数似然是这些项的负和（交叉熵）。

16. Worksheet - CNN Dimensions

English	Chinese (中文)
Question: Calculate output size for CNN: Image $N=28$, Filter $F=5$, Padding $P=2$, Stride $S=1$.	问题: 计算 CNN 输出尺寸：图像 $N=28$，卷积核 $F=5$，填充 $P=2$，步幅 $S=1$。
Solution: $O = \frac{N - F + 2P}{S} + 1 = \frac{28 - 5 + 4}{1} + 1 = 28$.	解答: $O = \frac{N - F + 2P}{S} + 1 = \frac{28 - 5 + 4}{1} + 1 = 28$。

17. Worksheet - PCA optimization

English	Chinese (中文)
Question: What is the optimization criterion for the first principal component u_1?	问题: 第一主成分 u_1 的优化标准是什么？
Solution: Maximize variance $u_1^T C u_1$ subject to $u_1^T u_1 = 1$. Equivalent to minimizing reconstruction error.	解答: 在约束 $u_1^T u_1 = 1$ 下最大化方差 $u_1^T C u_1$。等同于最小化重构误差。

18. Worksheet - K-Means Limitations

English	Chinese (中文)
Question: When does K-means clustering fail?	问题: K-means 聚类何时会失效？
Solution: It assumes spherical clusters with similar variance. Fails on moon-shapes, concentric circles, or clusters with widely different variances/sizes.	解答: 它假设聚类是球形的且方差相似。在月牙形、同心圆或方差/大小差异巨大的聚类上会失效。

19. Worksheet - Naive Bayes Assumption

English	Chinese (中文)
Question: What is the key assumption of Naive Bayes?	问题: 朴素贝叶斯 (Naive Bayes) 的关键假设是什么？
Solution: Features are conditionally independent given the class label.	解答: 在给定类别标签的情况下，特征是条件独立的。

20. Worksheet - ELBO

English	Chinese (中文)
Question: Why do we use the Evidence Lower Bound (ELBO)?	问题: 为什么我们要使用证据下界 (ELBO)？
Solution: Because the true evidence (marginal likelihood) is intractable. We approximate the posterior with a variational distribution. Maximizing ELBO minimizes the KL divergence between the approximate and true posterior.	解答: 因为真实的证据（边际似然）不仅难以计算。我们用变分分布来近似后验。最大化 ELBO 可以最小化近似后验与真实后验之间的 KL 散度。