

# Machine Learning Revision Summary (Comprehensive)

Covers material from Revision Worksheet Part 2 and Part 1 Solutions

## 1. Bernoulli Distribution (Q2.1)

Question (English)	题目 (中文)
Write the Bernoulli distribution equations for $P(Y=1)$ and $P(Y=0)$ , and combine them into a single equation $P(Y=y)$ .	写出 $P(Y=1)$ 和 $P(Y=0)$ 的伯努利分布方程，并将它们合并为一个方程 $P(Y=y)$ 。
Solution	解答
$P(Y=1) = \pi, P(Y=0) = 1-\pi.$ Combined: $P(Y=y) = \pi^y (1-\pi)^{1-y}.$	$P(Y=1) = \pi, P(Y=0) = 1-\pi.$ 合并后 : $P(Y=y) = \pi^y (1-\pi)^{1-y}.$

## 2. Log-Likelihood (Q2.4)

Question (English)	题目 (中文)
Given independent samples $x_i$ with outcomes $y_i$ , write the likelihood $L$ and the negative log-likelihood.	给定具有结果 $y_i$ 的独立样本 $x_i$ ，写出似然 $L$ 和负对数似然。
Solution	解答
Likelihood $L = \prod_{i=1}^N p(x_i; y_i).$ Negative log-likelihood $= -\sum_{i=1}^N \log p(x_i; y_i).$	似然 $L = \prod_{i=1}^N p(x_i; y_i).$ 负对数似然 $= -\sum_{i=1}^N \log p(x_i; y_i).$

## 3. Logistic Regression Optimization

Question (English)	题目 (中文)
Derive the derivative of the log-probability for logistic regression using $\pi(x_i) = \text{Sigmoid}(w^T x_i)$ .	使用 $\pi(x_i) = \text{Sigmoid}(w^T x_i)$ 推导逻辑回归对数概率的导数。
Solution	解答
Log-prob: $y_i \log \pi(x_i) + (1-y_i) \log(1-\pi(x_i))$ . Differentiating requires iterative methods as it's non-linear.	对数概率 : $y_i \log \pi(x_i) + (1-y_i) \log(1-\pi(x_i))$ 。由于是非线性的，微分求解需要迭代方法。

## 4. Auto-differentiation (Q3.2)

Question (English)	题目 (中文)
What are the adjoints for reverse mode vs forward mode auto-differentiation?	反向模式与前向模式自动微分的伴随变量 ( adjoints ) 分别是什么 ?
Solution	解答
Reverse mode: $\bar{v}_j = \partial y_j / \partial v_i$ (gradient of output wrt node). Forward mode: $\dot{v}_i = \partial v_i / \partial x_j$ (gradient wrt input).	反向模式 : $\bar{v}_j = \partial y_j / \partial v_i$ (输出相对于节点的梯度)。 前向模式 : $\dot{v}_i = \partial v_i / \partial x_j$ (相对于输入的梯度)。

## 5. Neural Networks - Hidden Layers (Q4.3)

Question (English)	题目 (中文)
Why do we use hidden layers in neural networks?	为什么我们在神经网络中使用隐藏层 ?
Solution	解答
They allow learning finer-grained features and creating non-linear decision boundaries. Downside: increased computation, interpretability issues, needs more data.	它们允许学习更细粒度的特征并创建非线性决策边界。缺点 : 计算量增加 , 可解释性问题 , 需要更多数据。

## 6. NN Dimensions (Q4.6)

Question (English)	题目 (中文)
For images of size 30x45 and 5 classes, what is the input size for the first layer (if flattened) and output size?	对于尺寸为 30x45 的图像和 5 个类别 , 第一层的输入尺寸 ( 如果展平 ) 和输出尺寸是多少 ?
Solution	解答
Input: $30 \times 45 = 1350$ . Output: 5 (one-hot encoding).	输入 : $30 \times 45 = 1350$ 。输出 : 5 ( 独热编码 )。

## 7. CNN Output Size (Q5.2)

Question (English)	题目 (中文)
Calculate CNN output size. Image N=28, Filter F=5, Padding P=2, Stride S=1. Also for S=2.	计算 CNN 输出尺寸。图像 N=28 , 滤波器 F=5 , 填充 P=2 , 步幅 S=1。以及 S=2 的情况。
Solution	解答
Formula: $O = (N - F + 2P)/S + 1$ For S=1: $O = (28 - 5 + 4)/1 + 1 = 28$ . For S=2: $O = (28 - 5 + 4)/2 + 1 = 14.5 \rightarrow 14$ .	公式 : $O = (N - F + 2P)/S + 1$ 对于 S=1 : $O = (28 - 5 + 4)/1 + 1 = 28$ 。 对于 S=2 : $O = (28 - 5 + 4)/2 + 1 = 14.5 \rightarrow 14$ 。

## 8. CNN Architecture (Q5.3)

Question (English)	题目 (中文)
Suggest a simple CNN architecture for 30x45 RGB images with 5 classes.	为具有 5 个类别的 30x45 RGB 图像建议一个简单的 CNN 架构。
Solution	解答
1. Conv ( $F=7, P=3, S=1$ ), 8 filters. 2. ReLU. 3. Conv ( $F=5, P=2, S=3$ ), 16 filters (reduces size). 4. ReLU. 5. FC layer (input 2400, out 600). 6. ReLU. 7. FC layer (out 5).	1. 卷积 ( $F=7, P=3, S=1$ ), 8个滤波器。 2. ReLU。 3. 卷积 ( $F=5, P=2, S=3$ ), 16个滤波器 (减小尺寸)。 4. ReLU。 5. 全连接层 (输入 2400, 输出 600)。 6. ReLU。 7. 全连接层 (输出 5)。

## 9. PCA Optimization (Q6.5)

Question (English)	题目 (中文)
What is the optimisation criterion for the first principal component $u_1$ ?	第一主成分 $u_1$ 的优化标准是什么？
Solution	解答
Maximise $u_1^T C u_1 + \lambda_1 (1 - u_1^T u_1)$ . Maximise variance subject to normalisation constraint. Equivalent to minimising reconstruction error.	最大化 $u_1^T C u_1 + \lambda_1 (1 - u_1^T u_1)$ 。在归一化约束下最大化方差。等同于最小化重构误差。

## 10. Auto-encoders (Q6.6 & Q6.7)

Question (English)	题目 (中文)
Describe the shape of an auto-encoder and how it is used for dimensionality reduction.	描述自动编码器的形状以及如何将其用于降维。
Solution	解答
Encoder reduces input size to a compressed latent space. Decoder reconstructs input. For dimensionality reduction, use only the encoder part.	编码器将输入尺寸减小到压缩的潜在空间。解码器重建输入。对于降维，仅使用编码器部分。

## 11. K-Means Limitations (Q7.3)

Question (English)	题目 (中文)
When does K-means clustering fail?	K-means 聚类何时失效？
Solution	解答
Fails when clusters are not spherical or have different spreads (variances). Examples: moon-shaped data, concentric circles.	当聚类不是球形或具有不同的分布 ( 方差 ) 时失效。例如：月牙形数据，同心圆。

## 12. Normalized Cut (Q7.5)

Question (English)	题目 (中文)
Define the Normalized Cut (Ncut) between sets A and B.	定义集合 A 和 B 之间的归一化割 (Normalized Cut)。
<b>Solution</b>  $Ncut(A,B) = Cut(A,B)/Vol(A) + Cut(A,B)/Vol(B).$ Where Cut is sum of weights between A and B, Vol is sum of degrees in the set.	<b>解答</b>  $Ncut(A,B) = Cut(A,B)/Vol(A) + Cut(A,B)/Vol(B).$ 其中 Cut 是 A 和 B 之间的权重之和，Vol 是集合中度的总和。

## 13. Naive Bayes (Q8.2)

Question (English)	题目 (中文)
What is the key assumption for Naive Bayes classifier?	朴素贝叶斯分类器的关键假设是什么？
<b>Solution</b>	<b>解答</b>
Features are conditionally independent.	特征是条件独立的。

## 14. ELBO (Q8.7)

Question (English)	题目 (中文)
Explain the Evidence Lower Bound (ELBO).	解释证据下界 (ELBO)。
<b>Solution</b>  Used because true evidence is intractable. We learn a variational distribution to approximate the posterior. The ELBO relates to the KL divergence between true and variational posteriors.	<b>解答</b>  因为真实的证据难以计算。我们要学习一个变分分布来近似后验。ELBO 与真实后验和变分后验之间的 KL 散度有关。

## 15. Basics - Regression vs Classification (Q2)

Question (English)	题目 (中文)
Give an example of a (1) regression and (2) classification problem.	举例说明 (1) 回归问题 和 (2) 分类问题。
<b>Solution</b>  Regression: Predicting global temp in 2050. Classification: Identifying isopod species from image.	<b>解答</b>  回归：预测2050年的全球气温。分类：从图像中识别等足虫物种。

## 16. Basics - Extrapolation vs Interpolation (Q6)

Question (English)	题目 (中文)
Give examples of Extrapolation and Interpolation.	举例说明外推和内插。
Solution	解答
Extrapolation: Weather forecasting. Interpolation: Predicting student grade from attendance within known class data.	外推：天气预报。内插：在已知班级数据范围内根据出勤率预测学生成绩。

## 17. Basics - Objective Function (Q7)

Question (English)	题目 (中文)
What is an Objective Function?	什么是目标函数？
Solution	解答
Function to be minimised or maximised (e.g., cost or loss function) for parameter estimation.	用于参数估计的需要最小化或最大化的函数（例如成本或损失函数）。

## 18. Basics - Circular Analysis (Q8)

Question (English)	题目 (中文)
Define Circular Analysis and give an example.	定义循环分析并举例。
Solution	解答
Selecting analysis details using test data. Example: Adjusting fMRI parameters to get 'best' result on test data.	使用测试数据选择分析细节。例如：调整 fMRI 参数以在测试数据上获得“最佳”结果。

## 19. Entropy Equation (Q2)

Question (English)	题目 (中文)
Write the equation for entropy of discrete random variable X.	写出离散随机变量 X 的熵方程。
Solution	解答
$H[X] = - \sum p(x_i) \log p(x_i).$	$H[X] = - \sum p(x_i) \log p(x_i).$

## 20. Conditional Entropy (Q6)

Question (English)	题目 (中文)
Define conditional entropy.	定义条件熵。
Solution	解答
Expectation (over Y) of the entropy of X given Y.	给定 Y 时 X 的熵的（关于 Y 的）期望。

## 21. Entropy Calculation (Q7)

Question (English)	题目 (中文)
Calculate conditional entropy of rain given cloud state. $P(\text{Cloud})=0.8$ . $P(\text{Rain} \text{Cloud})=0.3$ , $P(\text{Rain} \text{NoCloud})=0.05$ .	计算给定云状态下雨的条件熵。 $P(\text{Cloud})=0.8$ . $P(\text{Rain} \text{Cloud})=0.3$ , $P(\text{Rain} \text{NoCloud})=0.05$ .
Solution	解答
$H(\text{Rain} \text{Cloud})=0.88$ , $H(\text{Rain} \text{NoCloud})=0.29$ . Expected $H = 0.8 \times 0.88 + 0.2 \times 0.29 = 0.76$ bits.	$H(\text{Rain} \text{Cloud})=0.88$ , $H(\text{Rain} \text{NoCloud})=0.29$ . 期望 $H = 0.8 \times 0.88 + 0.2 \times 0.29 = 0.76$ bits.

## 22. Information Gain (Q8)

Question (English)	题目 (中文)
Define Information Gain.	定义信息增益。
Expected reduction in entropy given new information: $H(X) - H(X Y)$ .	给定新信息后的预期熵减少量： $H(X) - H(X Y)$ 。

## 23. Confusion Matrix (ML Q2)

Question (English)	题目 (中文)
Draw confusion matrix. 10 Bees (7 correct, 2 wasp, 1 fly). 5 Wasps (all correct). 8 Flies (4 correct, 4 wasp).	绘制混淆矩阵。10只蜜蜂（7只正确，2只黄蜂，1只苍蝇）。5只黄蜂（全部正确）。8只苍蝇（4只正确，4只黄蜂）。
Solution	解答
Rows(True)/Cols(Pred): Bee: [7,2,1] Wasp: [0,5,0] Fly: [0,4,4]	行(真实)/列(预测): 蜜蜂: [7,2,1] 黄蜂: [0,5,0] 苍蝇: [0,4,4]

## 24. Time Series Validation (ML Q3)

Question (English)	题目 (中文)
Traffic prediction using random cross-validation works well. (a) Problem? (b) Solution?	使用随机交叉验证的交通预测效果很好。(a) 问题？(b) 解决方案？
Solution	解答
(a) Neighbouring times highly correlated (data leakage). (b) Use time-series split (train past, test future) or different days.	(a) 相邻时间高度相关（数据泄漏）。 (b) 使用时间序列分割（训练过去，测试未来）或不同的日期。

## 25. Decision Trees - Purity (Q1)

Question (English)	题目 (中文)
How is 'purity' used in decision trees?	如何在决策树中使用“纯度”？
Solution	解答
Split data so child nodes are as pure as possible (unmixed labels / low variance).	分割数据，使子节点尽可能纯净（未混合的标签/低方差）。

## 26. Information Gain Calculation (DT Q2)

Question (English)	题目 (中文)
Calculate Information Gain for 'vibration' feature. 16 turbines, $H[\text{state}] = 0.95$ . Vib: 8 (5 fail, 3 ok). No Vib: 8 (1 fail, 7 ok).	计算“振动”特征的信息增益。16台涡轮机， $H[\text{state}] = 0.95$ 。 振动：8 (5故障, 3正常)。无振动：8 (1故障, 7正常)。
Solution	解答
$H[\text{state} \text{vib}] = 0.75$ . $IG = 0.95 - 0.75 = 0.2 \text{ bits}$ .	$H[\text{state} \text{vib}] = 0.75$ . $IG = 0.95 - 0.75 = 0.2 \text{ bits}$ .

## 27. Overfitting in Trees (DT Q3)

Question (English)	题目 (中文)
How to avoid overfitting in decision trees?	如何避免决策树中的过拟合？
Solution	解答
Pruning (pre or post).	剪枝 (预剪枝或后剪枝)。

## 28. Bagging Confidence Interval (Ens Q1)

Question (English)	题目 (中文)
How to use bagging for 95% CI on NN prediction?	如何使用 Bagging 进行神经网络预测的 95% 置信区间？
Solution	解答
Resample training data with replacement. Train models. Get distribution of predictions. Find 2.5% and 97.5% percentiles.	有放回地重采样训练数据。训练模型。获取预测分布。找到 2.5% 和 97.5% 分位数。

## 29. Random Forest Steps (Ens Q2)

Question (English)	题目 (中文)
Explain Random Forest algorithm.	解释随机森林算法。
Solution	解答
1. Bootstrap samples. 2. Build trees. 3. Subspace sampling (random feature subset) at each split. 4. Aggregate predictions.	1. Bootstrap 采样。 2. 构建树。 3. 每个分裂处的子空间采样 ( 随机特征子集 ) 。 4. 聚合预测。

## 30. Linear Regression SSE (LR Q1)

Question (English)	题目 (中文)
Write sum-squared error (SSE) in vector notation.	用向量符号写出平方误差和 (SSE)。
Solution	解答
$(y - Xw)^T(y - Xw)$ .	$(y - Xw)^T(y - Xw)$ 。

## 31. Mean Absolute Error (LR Q2)

Question (English)	题目 (中文)
Expression for Mean Absolute Error.	平均绝对误差的表达式。
Solution	解答
$1/N \sum  y_i - [X]_{i:} w $ .	$1/N \sum  y_i - [X]_{i:} w $ 。

## 32. Gradients (LR Q3)

Question (English)	题目 (中文)
Differentiate SSE and MAE wrt w.	对 SSE 和 MAE 关于 w 求导。
Solution	解答
SSE grad: $-2X^T(y - Xw)$ . MAE grad: $-1/N \sum [X]_{i:} \text{sgn}(y_i - [X]_{i:} w)$ .	SSE 梯度 : $-2X^T(y - Xw)$ 。 MAE 梯度 : $-1/N \sum [X]_{i:} \text{sgn}(y_i - [X]_{i:} w)$ 。

## 33. Design Matrix (LR Q5)

Question (English)	题目 (中文)
Design matrix for 3rd order polynomial with inputs 2, 3, 4, 6, 8.	输入为 2, 3, 4, 6, 8 的 3 阶多项式的设计矩阵。
Solution	解答
Rows are $[1, x, x^2, x^3]$ . E.g., row 1: $[1, 2, 4, 8]$ .	行是 $[1, x, x^2, x^3]$ 。 例如，第 1 行 : $[1, 2, 4, 8]$ 。

## 34. L2 Regularization (LR Q6)

Question (English)	题目 (中文)
What to add to cost function for L2 regularization?	为了 L2 正则化，需要在成本函数中添加什么？
Solution	解答
Add $\lambda w^T w$ .	添加 $\lambda w^T w$ 。

## 35. Gaussian Process Covariance (GP Q1)

Question (English)	题目 (中文)
Exponential kernel $k(x, x') = e^{- x-x' /10}$ . Train: (1,2), (3,4). Test: 2. Compute $k_{*f}$ and $K_{ff}$ .	指数核 $k(x, x') = e^{- x-x' /10}$ 。训练点 : (1,2), (3,4)。测试点 : 2。计算 $k_{*f}$ 和 $K_{ff}$ 。
Solution	解答
$k_{*f} \approx [0.90, 0.90]$ . $K_{ff} \approx [[1, 0.82], [0.82, 1]]$ . Posterior mean $\approx 2.99$ .	$k_{*f} \approx [0.90, 0.90]$ . $K_{ff} \approx [[1, 0.82], [0.82, 1]]$ . 后验均值 $\approx 2.99$ .

## 36. GP Definition (GP Q4)

Question (English)	题目 (中文)
Define Gaussian Process.	定义高斯过程。
Solution	解答
Stochastic process where every finite collection of random variables has a multivariate normal distribution.	随机过程，其中每个有限的随机变量集合都服从多元正态分布。

## 37. Uncertainty in Lengthscale (GP Q5)

Question (English)	题目 (中文)
How to handle uncertainty in lengthscale?	如何处理长度尺度的不确定性？
Solution	解答
Bayesian approach: Integrate out lengthscale using a hyperprior, or approximate by sampling from hyperprior.	贝叶斯方法：使用超先验积分掉长度尺度，或通过从超先验采样来近似。