

# Machine Learning Revision Summary (Comprehensive v3)

Covers Part 1 (Basics) and Part 2 (Advanced).

# PART 1: BASICS

## Supervised vs Unsupervised Learning (Basics Q1)

Question (English)	题目 (中文)
Define Supervised and Unsupervised Learning.	定义监督学习和无监督学习。
Solution	解答
Supervised Learning: Learning a mapping from inputs $x$ to outputs $y$ , given a labeled dataset of pairs $(x, y)$ . Example: Classification, Regression. Unsupervised Learning: Finding patterns or structure in data $x$ without any corresponding target labels $y$ . Example: Clustering, Dimensionality Reduction.	监督学习：在给定标记数据集 $(x, y)$ 的情况下，学习从输入 $x$ 到输出 $y$ 的映射。例如：分类，回归。 无监督学习：在没有对应目标标签 $y$ 的情况下，寻找数据 $x$ 中的模式或结构。例如：聚类，降维。

## Regression vs Classification (Basics 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。
Solution	解答
Regression: Predicting global temp in 2050 (continuous output). Classification: Identifying isopod species from image (discrete class output).	回归：预测2050年的全球气温（连续输出）。 分类：从图像中识别等足虫物种（离散类别输出）。

## Data Splits (Basics Q3)

Question (English)	题目 (中文)
Define Training set, Validation set, and Test set.	定义训练集、验证集和测试集。
Solution	解答
Training set: Used to fit the model parameters. Validation set: Used for model selection and hyperparameter tuning during training. Test set: Used only *once* at the end to evaluate the final performance on unseen data.	训练集：用于拟合模型参数。 验证集：在训练期间用于模型选择和超参数调整。 测试集：仅在最后使用*一次*，用于评估未见数据的最终性能。

## Leave-one-out Cross Validation (Basics Q4)

Question (English)	题目 (中文)
Describe Leave-one-out cross validation.	描述留一法交叉验证 (LOOCV)。
Solution	解答
An extreme case of k-fold CV where k equals the number of data points (N). We train on N-1 sample and test on 1 sample, repeating this N times. It is computationally expensive but uses almost all data for training.	k折交叉验证的极端情况，其中k等于数据点的数量(N)。我们在N-1个样本上训练并在1个样本上测试，重复N次。计算成本很高，但几乎使用了所有数据进行训练。

## Generalisation (Basics Q5)

Question (English)	题目 (中文)
What is Generalisation?	什么是泛化？
Solution	解答
The ability of a trained model to perform well on new, unseen data, not just the data it was trained on. Poor generalisation often implies overfitting.	训练好的模型在新的、未见过的数据上也能表现良好的能力，而不仅仅是在训练数据上。泛化能力差通常意味着过拟合。

## Extrapolation vs Interpolation (Basics 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: Predicting outside the range of training data, e.g., Weather forecasting into the future. Interpolation: Predicting within the range of known data, e.g., Predicting a student's grade based on attendance using relation learnt from the rest of the class.	外推：在训练数据范围之外进行预测，例如：未来的天气预报。 内插：在已知数据范围内进行预测，例如：根据从班级其他同学那里学到的关系，利用出勤率预测某个学生的成绩。

## Objective Function (Basics Q7)

Question (English)	题目 (中文)
What is an Objective Function?	什么是目标函数？
Solution	解答
A function to be minimised or maximised (e.g., cost function or loss function) to estimate the best parameters for a model.	用于估计模型最佳参数的需要最小化或最大化的函数（例如成本函数或损失函数）。

## Circular Analysis (Basics 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/parameters using the test data itself. Example: Adjusting fMRI preprocessing parameters to get the 'best' result on the test data. This invalidates the test result (data leakage).	使用测试数据本身选择分析细节/参数。 例子：调整 fMRI 预处理参数以在测试数据上获得“最佳”结果。这会使测试结果无效（数据泄漏）。

## Expectation Equation (Entropy Q1)

Question (English)	题目 (中文)
Write equation for expectation of a function $f(X)$ of discrete random variable $X$ .	写出离散随机变量 $X$ 的函数 $f(X)$ 的期望方程。
Solution	解答
$E[f(X)] = \sum p(x_i) f(x_i)$ .	$E[f(X)] = \sum p(x_i) f(x_i)$ 。

## Entropy Equation (Entropy 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
$H[X] = - \sum p(x_i) \log p(x_i)$ .	$H[X] = - \sum p(x_i) \log p(x_i)$ 。

## Log Base (Entropy Q3)

Question (English)	题目 (中文)
What base do you need to use for your log to get an answer in bits?	你需要使用什么底数的对数才能得到以比特 (bits) 为单位的答案？
Solution	解答
Base 2.	底数 2。

## Entropy of Dice (Entropy Q4)

Question (English)	题目 (中文)
What is the entropy of a fair dice roll, in bits?	掷公平骰子的熵是多少 (比特) ?
Solution	解答
The 6 outcomes have probability 1/6. $H = - \sum (1/6) \log_2(1/6) = - \log_2(1/6) = \log_2(6) \approx 2.585 \text{ bits.}$	6个结果的概率均为 1/6。 $H = - \sum (1/6) \log_2(1/6) = - \log_2(1/6) = \log_2(6) \approx 2.585 \text{ 比特。}$

## Entropy of Coin (Entropy Q5)

Question (English)	题目 (中文)
What is the entropy of an unfair coin toss (90% chance of heads)?	不公平抛硬币 (90% 几率正面向上) 的熵是多少 ?
Solution	解答
$P(H)=0.9, P(T)=0.1.$ $H = - (0.9 \log_2 0.9 + 0.1 \log_2 0.1) \approx 0.469 \text{ bits.}$	$P(H)=0.9, P(T)=0.1.$ $H = - (0.9 \log_2 0.9 + 0.1 \log_2 0.1) \approx 0.469 \text{ 比特。}$

## Conditional Entropy (Entropy Q6)

Question (English)	题目 (中文)
Define conditional entropy (in words).	定义条件熵 (用文字描述)。
Solution	解答
The expectation (over Y) of the entropy of X given Y. It represents the uncertainty remaining in X after observing Y.	给定 Y 时 X 的熵的 (关于 Y 的) 期望。它表示在观察到 Y 后 X 中剩余的不确定性。

## Entropy Calculation (Entropy 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
$H(\text{Rain} \text{Cloud}) (p=0.3) = 0.88$ $H(\text{Rain} \text{NoCloud}) (p=0.05) = 0.29$ $H(\text{Rain} C) = P(C)H(R C) + P(\sim C)H(R \sim C)$ $= 0.8 \times 0.88 + 0.2 \times 0.29 = 0.76 \text{ bits.}$	$H(\text{Rain} \text{Cloud}) (p=0.3) = 0.88$ $H(\text{Rain} \text{NoCloud}) (p=0.05) = 0.29$ $H(\text{Rain} C) = P(C)H(R C) + P(\sim C)H(R \sim C)$ $= 0.8 \times 0.88 + 0.2 \times 0.29 = 0.76 \text{ 比特。}$

## Information Gain (Entropy Q8)

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

## Marginal Probability (Entropy Q9)

Question (English)	题目 (中文)
Using example in Q7, what is the marginal probability of rain?	使用 Q7 中的示例，下雨的边缘概率是多少？
Solution	解答
$P(\text{Rain}) = P(R C)P(C) + P(R \sim C)P(\sim C)$ = $0.3 \times 0.8 + 0.05 \times 0.2$ = $0.24 + 0.01 = 0.25$ (25%).	$P(\text{Rain}) = P(R C)P(C) + P(R \sim C)P(\sim C)$ = $0.3 \times 0.8 + 0.05 \times 0.2$ = $0.24 + 0.01 = 0.25$ (25%)。

## Entropy Comparison (Entropy Q10)

Question (English)	题目 (中文)
What is the entropy of rain/no-rain? Confirm it is not less than the conditional entropy.	下雨/不下雨的熵是多少？确认它小于条件熵。
Solution	解答
Using $P(\text{Rain})=0.25$ : $H(R) = -(0.25 \log_2 0.25 + 0.75 \log_2 0.75) \approx 0.81$ bits. 0.81 (Unconditional) $\geq$ 0.76 (Conditional in Q7). Holds true.	使用 $P(\text{Rain})=0.25$ : $H(R) = -(0.25 \log_2 0.25 + 0.75 \log_2 0.75) \approx 0.81$ 比特。 0.81 (无条件) $\geq$ 0.76 (Q7中的条件熵)。成立。

## Performance Metrics (End-to-End Q1)

Question (English)	题目 (中文)
How to compute: (a) RMSE, (b) Mean Absolute Error, (c) Negative Log-Predictive Density (NLPD).	如何计算：(a) 均方根误差 (RMSE) , (b) 平均绝对误差 , (c) 负对数预测密度 (NLPD)。
Solution	解答
(a) RMSE: $\sqrt{\frac{1}{N} \sum (y - y_{\text{pred}})^2}$ (b) MAE: $\frac{1}{N} \sum  y - y_{\text{pred}} $ (c) NLPD: $-\frac{1}{N} \sum \log p(y_i   x_i)$ (Evaluates probabilistic forecasts)	(a) RMSE: $\sqrt{\frac{1}{N} \sum (y - y_{\text{pred}})^2}$ (b) MAE: $\frac{1}{N} \sum  y - y_{\text{pred}} $ (c) NLPD: $-\frac{1}{N} \sum \log p(y_i   x_i)$ (评估概率预测)

## Confusion Matrix (End-to-End 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]

## Time Series Validation (End-to-End 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) Data Leakage: Neighbouring times are highly correlated. Random split allows training on t+1 to predict t. (b) Use Time-Series Split (Train on past, Test on future) or split by different days to ensure independence.	(a) 数据泄漏：相邻时间高度相关。随机分割允许利用t+1的数据训练来预测t。 (b) 使用时间序列分割（在过去的 data 上训练，在未来的 data 上测试）或按不同的日期分割以确保独立性。

## Decision Trees - Purity (DT 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
We choose split features to maximize purity of child nodes. Purity means labels are unmixed (low entropy/Gini for classification) or low variance (for regression).	我们选择分裂特征以最大化子节点的纯度。纯度意味着标签未混合（分类中低熵/基尼系数）或低方差（回归中）。

## Information Gain Calculation (DT Q2)

Question (English)	题目 (中文)
Calculate Information Gain for 'vibration' feature. 16 turbines, H[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[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
$H[state vib] = 0.75.$ $IG = 0.95 - 0.75 = 0.2 \text{ bits.}$	$H[state vib] = 0.75.$ $IG = 0.95 - 0.75 = 0.2 \text{ 比特。}$

## 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-pruning: limit depth; or post-pruning: remove nodes).	剪枝 ( 预剪枝 : 限制深度 ; 或后剪枝 : 移除节点 ) 。

## Bagging Confidence Interval (Ensemble 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
1. Resample training data with replacement (bootstrap). 2. Train multiple models. 3. Get distribution of predictions. 4. Find 2.5% and 97.5% percentiles of the predictions.	1. 有放回地重采样训练数据 (Bootstrap)。 2. 训练多个模型。 3. 获取预测分布。 4. 找到预测值的 2.5% 和 97.5% 分位数。

## Random Forest Steps (Ensemble Q2)

Question (English)	题目 (中文)
Explain Random Forest algorithm.	解释随机森林算法。
Solution	解答
1. Bootstrap: Generate samples with replacement. 2. Build Trees: Train decision trees on samples. 3. Subspace Sampling: At each split, consider only a random subset of features. 4. Aggregate: Majority vote (classification) or average (regression).	1. Bootstrap : 有放回地生成样本。 2. 构建树 : 在样本上训练决策树。 3. 子空间采样 : 在每个分裂处 , 仅考虑随机的特征子集。 4. 聚合 : 多数投票 ( 分类 ) 或平均 ( 回归 ) 。

## 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
$(y - Xw)^T(y - Xw)$ .	$(y - Xw)^T(y - Xw)$ 。

## 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 $ 。

## 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: $-1/N \sum [X]_i \operatorname{sgn}(y_i - [X]_i w)$ .	SSE 梯度 : $-2X^T(y - Xw)$ 。 MAE 梯度 : $-1/N \sum [X]_i \operatorname{sgn}(y_i - [X]_i w)$ 。

## Likelihood vs SSE (LR Q4)

Question (English)	题目 (中文)
Show maximizing likelihood with Gaussian noise is equivalent to minimizing SSE.	证明在具有高斯噪声的情况下最大化似然等同于最小化 SSE。
Solution	解答
Gaussian Likelihood includes term $\exp(-\text{SSE})$ . Maximizing log-likelihood involves maximizing $-\text{SSE}$ , which is equivalent to Minimizing SSE.	高斯似然包含 $\exp(-\text{SSE})$ 项。最大化对数似然涉及最大化 $-\text{SSE}$ ，这等同于最小化 SSE。

## 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 : $[1, 2, 4, 8]$ 。

## 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 $\lambda w^T w$ (penalty on squared weights).	添加 $\lambda w^T w$ ( 对权重平方的惩罚 )。

## L1 Regularization (LR Q7)

Question (English)	题目 (中文)
What to add for L1 regularization?	为了 L1 正则化，需要添加什么？
Solution	解答
Add $\lambda \sum  w_i $ (penalty on absolute weights). Promotes sparsity.	添加 $\lambda \sum  w_i $ ( 对权重绝对值的惩罚 )。促进稀疏性。

## 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
$k_{*f} = [k(2,1), k(2,3)] \approx [0.90, 0.90]$ . $K_{ff} = [[k(1,1), k(1,3)], [k(3,1), k(3,3)]] \approx [[1, 0.82], [0.82, 1]]$ . Posterior mean = $k_{*f} K^{-1} y \approx 2.99$ .	$k_{*f} = [k(2,1), k(2,3)] \approx [0.90, 0.90]$ . $K_{ff} = [[k(1,1), k(1,3)], [k(3,1), k(3,3)]] \approx [[1, 0.82], [0.82, 1]]$ . 后验均值 = $k_{*f} K^{-1} y \approx 2.99$ .

## GP Definition (GP Q4)

Question (English)	题目 (中文)
Define Gaussian Process.	定义高斯过程。
Solution	解答
A stochastic process where every finite collection of random variables has a multivariate normal distribution.	一种随机过程，其中每个有限的随机变量集合都服从多元正态分布。

## Uncertainty in Lengthscale (GP Q5)

Question (English)	题目 (中文)
How to handle uncertainty in lengthscales?	如何处理长度尺度的不确定性？
Solution	解答
Bayesian approach: Place a hyperprior on the lengthscale and integrate it out (marginalize), or approximate by sampling from the hyperprior and averaging predictions.	贝叶斯方法：如果对长度尺度设置超先验并将其积分掉（边缘化），或者通过从超先验采样并平均预测结果来近似。

## Kernel Trick (Kernel Q1)

Question (English)	题目 (中文)
Review shape of $XX^T$ vs $X^TX$ . Which is easier to invert?	回顾 $XX^T$ 与 $X^TX$ 的形状。哪一个更容易求逆？
Solution	解答
If $N$ (data points) > $D$ (features), $X^TX$ ( $D \times D$ ) is smaller/easier. If $D \gg N$ , $XX^T$ ( $N \times N$ ) is smaller. Kernel trick uses $N \times N$ Gram matrix.	如果 $N$ (数据点) > $D$ (特征)，则 $X^TX$ ( $D \times D$ ) 更小/更容易。如果 $D \gg N$ ，则 $XX^T$ ( $N \times N$ ) 更小。核技巧使用 $N \times N$ Gram 矩阵。

## PART 2: ADVANCED

### 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 : $P(Y=y) = \pi^y (1-\pi)^{1-y}.$

### 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 $L = \prod p(x_i; y_i).$ Negative log-likelihood $= -\sum \log p(x_i; y_i).$	似然 $L = \prod p(x_i; y_i).$ 负对数似然 $= -\sum \log p(x_i; y_i).$

### Logistic Regression Optimization

Question (English)	题目 (中文)
Derive the derivative of the log-probability for logistic regression using $\pi(x_i) = \text{Sigmoid}(w^\top x_i)$ .	使用 $\pi(x_i) = \text{Sigmoid}(w^\top x_i)$ 推导逻辑回归对数概率的导数。
Solution	解答
Log-prob involves $y_i \log \pi + (1-y_i) \log(1-\pi)$ . Sigmoid derivative is $\pi(1-\pi)$ . Result is complex but solvable with gradient descent (iterative).	对数概率涉及 $y_i \log \pi + (1-y_i) \log(1-\pi)$ 。Sigmoid 导数是 $\pi(1-\pi)$ 。 结果复杂，但可用梯度下降（迭代）求解。

### 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: Adjoint $\bar{v}_i = \partial \text{Output} / \partial v_i$ (Propagates gradient backwards). Forward mode: Adjoint $\dot{v}_i = \partial v_i / \partial \text{Input}$ (Propagates derivatives forwards).	反向模式：伴随变量 $\bar{v}_i = \partial \text{输出} / \partial v_i$ (向后传播梯度)。 前向模式：伴随变量 $\dot{v}_i = \partial v_i / \partial \text{输入}$ (向前传播导数)。

## 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 non-linear decision boundaries and hierarchical features. Without them, NN is just a linear model.	它们允许学习非线性决策边界和分层特征。如果没有它们，神经网络就只是一个线性模型。

## 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: $30 \times 45 = 1350$ units. Output: 5 units (one-hot encoding).	输入 : $30 \times 45 = 1350$ 个单元。 输出 : 5 个单元 (独热编码)。

## 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
$O = (N - F + 2P)/S + 1.$ $S=1: (28-5+4)/1 + 1 = 28.$ $S=2: (28-5+4)/2 + 1 = 14.5 \rightarrow 14$ (floor).	$O = (N - F + 2P)/S + 1.$ $S=1: (28-5+4)/1 + 1 = 28.$ $S=2: (28-5+4)/2 + 1 = 14.5 \rightarrow 14$ (向下取整)。

## 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
Input(30x45x3) -> Conv(F=7) -> ReLU -> Conv(F=5, S=3) -> ReLU -> FC(600) -> Output(5).	输入(30x45x3) -> 卷积(F=7) -> ReLU -> 卷积(F=5, S=3) -> ReLU -> 全连接(600) -> 输出(5)。

## 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
Maximize variance: $u_1^T C u_1$ subject to $\ u_1\ =1$ . Equivalent to minimising reconstruction error.	最大化方差 : $u_1^T C u_1$ , 约束条件为 $\ u_1\ =1$ 。 等同于最小化重构误差。

## Auto-encoders (Q6.6/6.7)

Question (English)	题目 (中文)
Describe Auto-encoder shape and use for dimensionality reduction.	描述自动编码器的形状及其在降维中的用途。
Solution	解答
Hourglass shape: Encoder shrinks input to latent space (bottleneck); Decoder expands back. For reduction, use only the Encoder.	沙漏形状：编码器将输入缩小到潜在空间（瓶颈）；解码器将其展开。对于降维，仅使用编码器。

## K-Means Failures (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
Non-spherical clusters (e.g., moons), different variances/densities, or significant outliers.	非球形聚类（例如月牙形），不同的方差/密度，或显著的异常值。

## Normalized Cut (Q7.5)

Question (English)	题目 (中文)
Define Normalized Cut (Ncut) between sets A and B.	定义集合 A 和 B 之间的归一化割 (Normalized Cut)。
Solution	解答
$Ncut(A,B) = Cut(A,B)/Vol(A) + Cut(A,B)/Vol(B)$ . Normalized by volume (total degrees), encouraging balanced clusters.	$Ncut(A,B) = Cut(A,B)/Vol(A) + Cut(A,B)/Vol(B)$ 。 按体积 (总度数) 归一化，鼓励平衡的聚类。

## Naive Bayes Assumption (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？
Solution	解答
Features are conditionally independent given the class label.	在给定类别标签的情况下，特征是条件独立的。

## ELBO (Q8.7)

Question (English)	题目 (中文)
Explain the Evidence Lower Bound (ELBO).	解释证据下界 (ELBO)。
Solution	解答
A lower bound on the log-evidence ( $\log p(x)$ ). Maximizing ELBO minimizes KL divergence between approximate posterior and true posterior.	对数证据 ( $\log p(x)$ ) 的下界。最大化 ELBO 可最小化近似后验与真实后验之间的 KL 散度。