Okay, here is a comprehensive learning note summarizing the key concepts from the PDF document, including both Chinese and English explanations as per the original text:

I. Overview of Machine Learning (机器学习)

1. **Two Perspectives on Machine Learning (ML)**
   - **Data-first view:** ML is a set of tools for extracting knowledge from data. (数据为先的视角：机器学习是一套从数据中提取知识的工具。)
   - **Deployment-first view:** ML is a set of tools together with a methodology for solving problems using data. (部署为先的视角：机器学习是一套工具，以及一套使用数据解决问题的方法论。)
2. **Core Concepts**
   - **Dataset (数据集):** A collection of items described by a set of attributes. (由一系列具有相同属性的条目组成。)
   - **Model (模型):** A representation of knowledge. (知识的表示。)
   - **Taxonomy (分类体系):** A system for classifying different types of ML problems, techniques, and models. (用于区分不同类型的机器学习问题、技术和模型。)
3. **Machine Learning Taxonomy**

```
Machine Learning（机器学习）
├── Supervised Learning（监督学习）
│   ├── Classification（分类）
│   └── Regression（回归）
└── Unsupervised Learning（无监督学习）
    ├── Density Estimation（密度估计）
    └── Structure Analysis（结构分析）
```

II. Regression (回归)

1. **Definition**
   - Regression is a supervised problem: Our goal is to predict the value of one attribute (label) using the remaining attributes (predictors). (回归是一个监督学习问题：我们的目标是使用其余属性（预测变量）来预测某个属性的值（标签）。)
   - The label is a continuous variable. (标签是一个连续变量。)
   - Our job is then to find the best model that assigns a unique label to a given set of predictors. (我们的任务是找到最佳模型，该模型为给定的一组预测变量分配唯一的标签。)
   - We use datasets consisting of labelled samples. (我们使用由标记样本组成的数据集。)
2. **Problem Formulation (问题形式化)**
   - **Predictors (预测变量):** Input attributes. (输入属性。)
   - **Label (标签):** The output attribute to be predicted, a continuous value. (需要预测的输出属性，是一个连续值。)
   - **Model (模型):** A function that maps predictors to a label. (将预测变量映射到标签的函数。)
   - **Goal (目标):** Find the best model that accurately predicts the label given the predictors. (找到在给定预测变量的情况下准确预测标签的最佳模型。)
3. **Examples of Regression Problems (回归问题的例子)**
   - Predict the energy consumption of a household, given the location of the house, household size, income, intensity of occupation. (在给定房屋位置、家庭规模、收入、入住强度的情况下，预测家庭的能源消耗。)
   - Predict future values of a company stock, given past stock prices. (在给定过去股价的情况下，预测公司股票的未来价值。)
   - Predict distance driven by a vehicle given its speed and journey duration. (在给定车辆速度和行驶时间的情况下，预测车辆行驶的距离。)
   - Predict demand given past demand and currency exchange rate. (在给定过去的需求和汇率的情况下，预测需求。)
   - Predict tomorrow's temperature given today's temperature and pressure. (在给定今天的温度和气压的情况下，预测明天的温度。)
   - Predict the probability to develop a specific heart condition given BMI, alcohol consumption, diet, number of daily steps. (在给定 BMI、酒精摄入量、饮食、每日步数的情况下，预测患特定心脏病的概率。)
4. **Association and Causation (关联与因果)**
   - Prediction models are sometimes interpreted through a causal lens: the predictor is the cause, the label its effect. However this is not correct. (预测模型有时会被从因果关系的角度解读：预测变量是因，标签是果。然而，这是不正确的。)
   - Our ability to build predictors is due to association between attributes, rather than causation. (我们构建预测模型的能力源于属性之间的关联，而不是因果关系。)
   - **Take-home message:** In machine learning we don't build causal models! (关键信息：在机器学习中，我们不构建因果模型！)
5. **Mathematical Notation (数学符号)**
   - **Population (总体):**
     - x: Predictor attribute (预测变量属性)

- y: Label attribute (标签属性)
- **Dataset (数据集):**
  - N: Number of samples (样本数量), i identifies each sample (标识每个样本)
  - x_i: Predictor of sample i (样本 i 的预测变量)
  - y_i: Actual label of sample i (样本 i 的实际标签)
  - (x_i, y_i): Sample i (样本 i), $\{(x_i, y_i) : 1 \le i \le N\}$: Entire dataset (整个数据集)
- **Model (模型):**
  - $f(\cdot)$: Denotes the model (表示模型)
  - $\hat{y}_i = f(x_i)$: Predicted label for sample i (样本 i 的预测标签)
  - $y_i - \hat{y}_i$: Prediction error for sample i (样本 i 的预测误差)

6. **Model Quality (模型质量)**
   - **Squared Error (平方误差):** $e^2 = (y_i - \hat{y}_i)^2$ a common quantity used in regression to encapsulate the notion of single prediction quality. (回归中用于概括单个预测质量概念的常用量。)
   - **Sum of Squared Errors (SSE):** $E_{SSE} = e^2 + e^2 + \ldots + e^2 = \Sigma e^2$ (误差平方和)
   - **Mean Squared Error (MSE):** $E_{MSE} = (1/N) \Sigma e^2$ (均方误差)
7. **Regression as an Optimization Problem (作为优化问题的回归)**
   - Goal: Find the model with the lowest E_{MSE}. (目标：找到 E_{MSE} 最低的模型。)
   - $f_{best}(x) = \arg\min (1/N) \Sigma (y_i - f(x_i))^2$
   - Finding such a model is an optimisation problem. (找到这样的模型是一个优化问题。)
   - Note that we are defining regression as finding the model that minimises EMSE on the dataset, without considering what happens once deployed. (请注意，我们将回归定义为在数据集上找到最小化 E_{MSE} 的模型，而不考虑部署后发生的情况。)

III. Basic Regression Models (基本回归模型)

1. **Our Regression Learner (我们的回归学习器)**
   - **Priors (先验):** Type of model (linear, polynomial, etc.). (模型的类型（线性、多项式等）。)
   - **Data (数据):** Labelled samples (predictors and true label). (标记样本（预测变量和真实标签）。)
   - **Model (模型):** Predicts a label based on the predictors. (根据预测变量预测标签。)
2. **Simple Regression (简单回归)**
   - Simple regression considers one predictor x and one label y. (简单回归考虑一个预测变量 x 和一个标签 y。)
3. **Simple Linear Regression (简单线性回归)**
   - Model: $f(x) = w_0 + w_1x$
   - Predicted label: $\hat{y}_i = f(x_i) = w_0 + w_1x_i$
   - **Parameters (参数):** w_0 (intercept, 截距), w_1 (gradient, 斜率)
   - **Training (训练):** Tuning the parameters using a dataset. (使用数据集调整参数。) Also referred to as fitting the model to the training dataset. (也称为将模型拟合到训练数据集。)
4. **Beyond Linearity (超越线性)**
   - Considering models beyond simple linear functions. (考虑简单线性函数之外的模型。)
5. **Simple Polynomial Regression (简单多项式回归)**
   - Model: $f(x_i) = w_0 + w_1x_i + w_2x^2 + \ldots + w_Dx^D$
   - D: Degree of the polynomial (多项式的次数), a hyperparameter (超参数)
   - **Hyperparameter (超参数):** A parameter that defines the family of models. (定义模型系列的参数。) Setting its value results in a different family, with a different collection of parameters. (设置超参数的值会得到不同的模型系列，具有不同的参数集合。)
   - $D = 1$: Linear family (线性系列)
   - $D = 2$: Quadratic family (二次系列)
   - $D = 3$: Cubic family (三次系列)

IV. Multiple Regression (多元回归)

1. **Definition:** In multiple regression there are two or more predictors. Given item i, we will denote each individual predictor as $x_{i,1}, x_{i,2}, \ldots, x_{i,K}$, where K is the number of predictors. (在多元回归中，有两个或更多的预测变量。给定条目 i，我们将每个单独的预测变量表示为 $x_{i,1}, x_{i,2}, \ldots, x_{i,K}$，其中 K 是预测变量的数量。)
2. **Vector Notation (向量表示):**
   - $x_i = [1, x_{i,1}, x_{i,2}, \ldots, x_{i,K}]^T$

    - `xi,k`: The k-th predictor of the i-th sample. (第 `i` 个样本的第 `k` 个预测变量。)
    - The constant 1 is prepended for convenience. (为方便起见，添加了常数 `1`。)
    - `yi = f(xi)`
3. **Multiple Linear Regression Formulation (多元线性回归公式):**
    - `f(xi) = w^T xi = w0 + w1xi,1 + ... + wKxi,K`
    - `w = [w0, w1, ..., wK]^T`: Model's parameter vector (模型的参数向量)
4. **Design Matrix (设计矩阵) and Label Vector (标签向量):**
    - **Design Matrix X:**

      ```
      X = [x1, ..., xN]^T =  [1  x1,1  x1,2  ... x1,K]
                            [1  x2,1  x2,2  ... x2,K]
                            [ :   :     :        :  ]
                            [1  xN,1  xN,2  ... xN,K]
      ```

    - **Label Vector y:**

      ```
      y = [y1, ..., yN]^T = [y1]
                           [y2]
                           [ :]
                           [yN]
      ```

5. **The Least Squares Solution (最小二乘解):**
    - `w_best = (X^T X)^-1 X^T y`
    - This is an exact or analytical solution and is known as the least squares solution. It is valid for simple and multiple linear regression. (这是一个精确或解析解，被称为最小二乘解。它对简单和多元线性回归都有效。)
    - This solution can also be used for polynomial models, by treating the powers of the predictor as predictors themselves. (通过将预测变量的幂视为预测变量本身，此解决方案也可用于多项式模型。)
    - Note that the inverse matrix `(X^T X)^-1` exists when all the columns in `X` are linearly independent. (请注意，当 `X` 中的所有列都线性无关时，逆矩阵 `(X^T X)^-1` 存在。)

V. Other Models for Regression (其他回归模型)

- **Exponential (指数模型)**
- **Sinusoids (正弦曲线模型)**
- **Radial basis functions (径向基函数模型)**
- **Splines (样条函数模型)**
- And many more!

The mathematical formulation is identical and only the expression for `f(·)` changes. (数学公式是相同的，只有 `f(·)` 的表达式会发生变化。)

VI. Logistic Model (逻辑模型)

- Used when the label represents a proportion or a probability, i.e. a quantity between 0 and 1. (当标签表示比例或概率时使用，即 0 到 1 之间的数量。)
- The logistic function `p(d)` is defined as `p(d) = e^d / (1 + e^d) = 1 / (1 + e^-d)`. (逻辑函数 `p(d)` 定义为 `p(d) = e^d / (1 + e^d) = 1 / (1 + e^-d)`。)
- By using `d = w0 + w1x`, we can translate the logistic function and change its gradient. (通过使用 `d = w0 + w1x`，我们可以平移逻辑函数并改变其梯度。)

VII. Other Quality Metrics (其他质量指标)

1. **Root Mean Squared Error (RMSE):** $ERMSE = \sqrt{((1/N)\ \Sigma\ e^2)}$. Measures the sample standard deviation of the prediction error. (均方根误差：衡量预测误差的样本标准差。)
2. **Mean Absolute Error (MAE):** `EMAE = (1/N) Σ |ei|`. Measures the average of the absolute prediction error. (平均绝对误差：衡量绝对预测误差的平均值。)
3. **R-squared (R^2):** $ER = 1 - (\Sigma\ e^2 / \Sigma\ (yi - \bar{y})^2)$, where $\bar{y} = (1/N)\ \Sigma\ yi$. Measures the proportion of the variance in the response that is predictable from the predictors. (R^2：衡量响应中可由预测变量预测的方差比例。)

VIII. Flexibility, Interpretability, and Generalisation (灵活性、可解释性和泛化性)

1. **Flexibility (灵活性)**
    - Models allow us to generate multiple shapes by tuning their parameters. (模型允许我们通过调整其参数来生成多种形状。)
    - We talk about the degrees of freedom or the complexity of a model to describe its ability to generate different shapes, i.e. its flexibility. (我们讨论模型的自由度或复杂性，以描述其生成不同形状的能力，即其灵活性。)

    - The degrees of freedom of a model are in general related to the number of parameters of the model: (模型的自由度通常与其参数数量有关：)
        - A linear model `y = w0 + w1x` has two parameters and is inflexible, as it can only generate straight lines. (线性模型 `y = w0 + w1x` 有两个参数，并且不灵活，因为它只能生成直线。)
        - A cubic model `y = w0 + w1x + w2x² + w3x³` has 4 parameters and is more flexible than a linear one. (三次模型 `y = w0 + w1x + w2x² + w3x³` 有 4 个参数，并且比线性模型更灵活。)
    - The flexibility of a model is related to its interpretability and accuracy and there is a trade-off between the two. (模型的灵活性与其可解释性和准确性有关，并且两者之间存在权衡。)
2. **Interpretability (可解释性)**
    - Model interpretability is crucial for us, as humans, to understand in a qualitative manner how a predictor is mapped to a label. (模型的可解释性对于我们人类来说至关重要，可以定性地理解预测变量如何映射到标签。)
    - Inflexible models produce solutions that are usually simpler and easier to interpret. (不灵活的模型产生的解决方案通常更简单，更容易解释。)
3. **Quality on the Training Dataset (训练数据集上的质量)**
    - The quality of a model on a training dataset is also related to its flexibility. During training, the error produced by flexible models is in general lower. (模型在训练数据集上的质量也与其灵活性有关。在训练期间，灵活模型产生的误差通常较低。)
4. **Generalisation (泛化性)**
    - We have considered the training MSE, i.e. the quality of regression models on the training dataset. (我们已经考虑了训练均方误差，即回归模型在训练数据集上的质量。)
    - **Generalisation:** The ability of our model to successfully translate what we was learnt during the learning stage to deployment. (泛化性：我们的模型将学习阶段学到的知识成功转化到部署阶段的能力。)
    - **Deployment MSE:** The MSE of a model when presented with new data. (部署均方误差：模型在遇到新数据时的均方误差。)
5. **Underfitting and Overfitting (欠拟合与过拟合)**
    - By comparing the performance of models during training and deployment, we can observe three different behaviours: (通过比较模型在训练和部署期间的性能，我们可以观察到三种不同的行为：)
        - **Underfitting (欠拟合):** Large training and deployment errors are produced. The model is unable to capture the underlying pattern. Rigid models lead to underfitting. (产生较大的训练和部署误差。该模型无法捕捉到潜在的模式。刚性模型会导致欠拟合。)
        - **Overfitting (过拟合):** Small errors are produced during training, large errors during deployment. The model is memorising irrelevant details. Too complex models and not enough data lead to overfitting. (训练期间产生较小的误差，部署期间产生较大的误差。该模型记住了不相关的细节。过于复杂的模型和数据不足会导致过拟合。)
        - **Just right (恰到好处):** Low training and deployment errors. The model is capable of reproducing the underlying pattern and ignores irrelevant details. (训练和部署误差都很低。该模型能够再现潜在的模式并忽略不相关的细节。)
    - **Remember this:** Generalisation can only be assessed by comparing training and deployment performance, not by just looking at how each model fits the training data. (请记住：泛化性只能通过比较训练和部署性能来评估，而不仅仅是通过查看每个模型如何拟合训练数据。)

IX. Final Historical Note (最后的历史注记)

- The term "regression" comes from the observation by Galton in the 19th century that children of tall people tend to be taller than average, but not as tall as their parents. This phenomenon is called **regression to the mean** (回归均值).

This comprehensive note covers the essential concepts and definitions from the provided PDF document on regression in machine learning. Remember to review the examples and diagrams in the original document for a better visual understanding. Please let me know if you have any other questions.