Supervised Learning: The output can be continuous (regression) or discrete (classification). Supervised learning means, you are told the correct answer so you can learn from that.

Unsupervised Learning: Clustering algorithms would be included in this.

Gradient Descent In Practice: Make sure features are on a similar scale to ensure speedy convergence. Get every feature into approximately -1<x<1 range.

Normal Equation vs. Gradient Descent: Normal equation is more efficient when number of features is small. When number of features is greater than 10,000 matrix inversion might be costly.