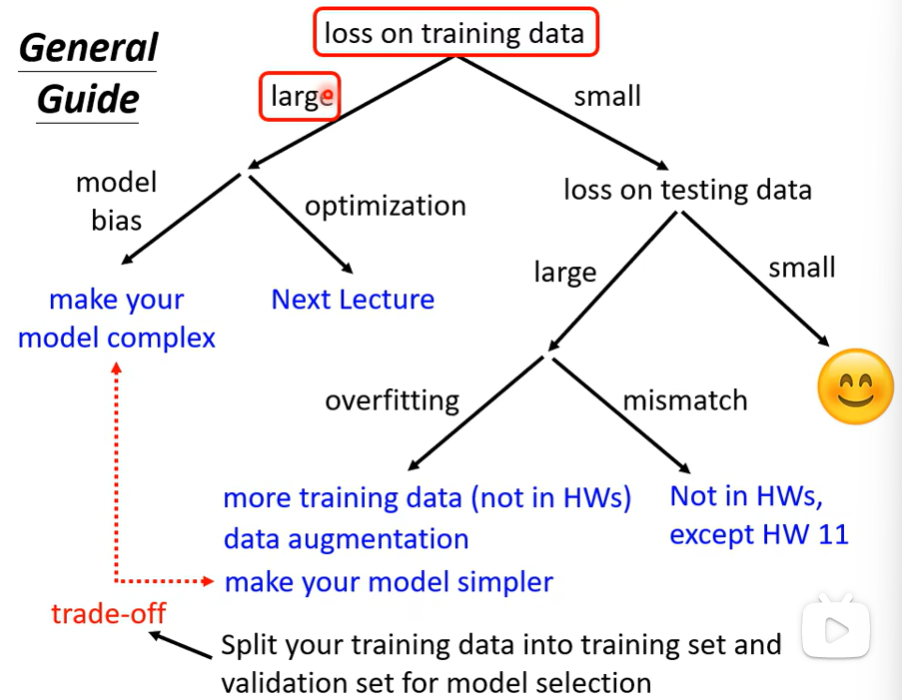
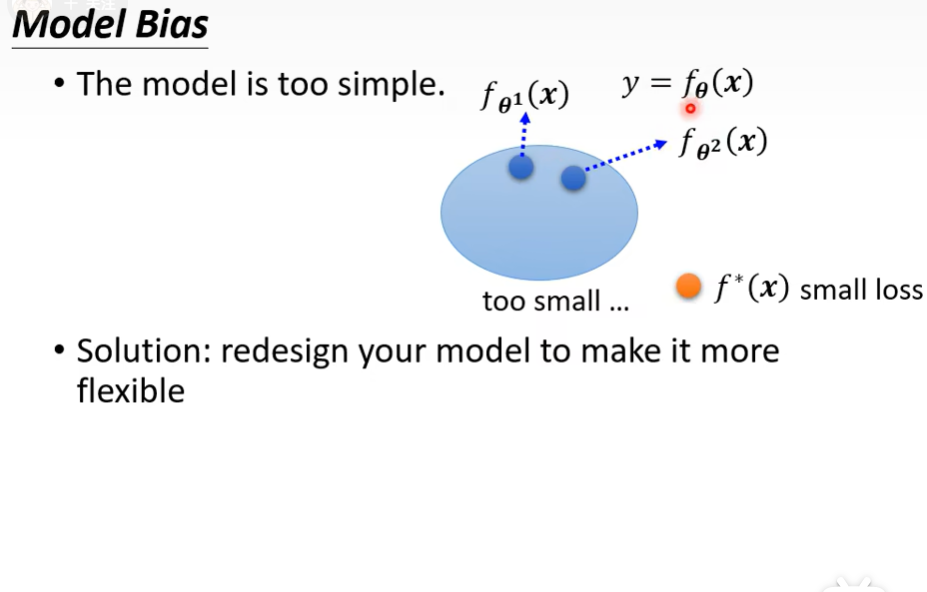
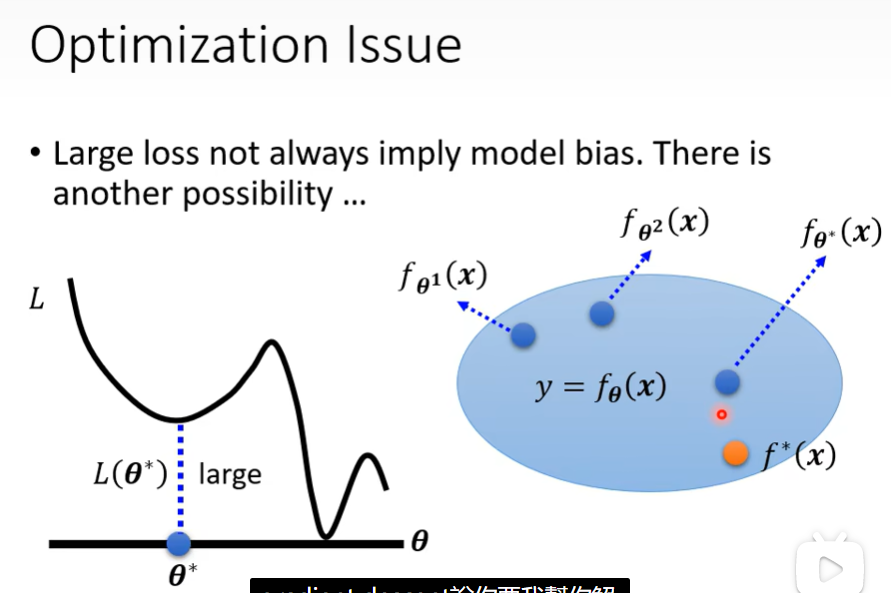
**General Guide**

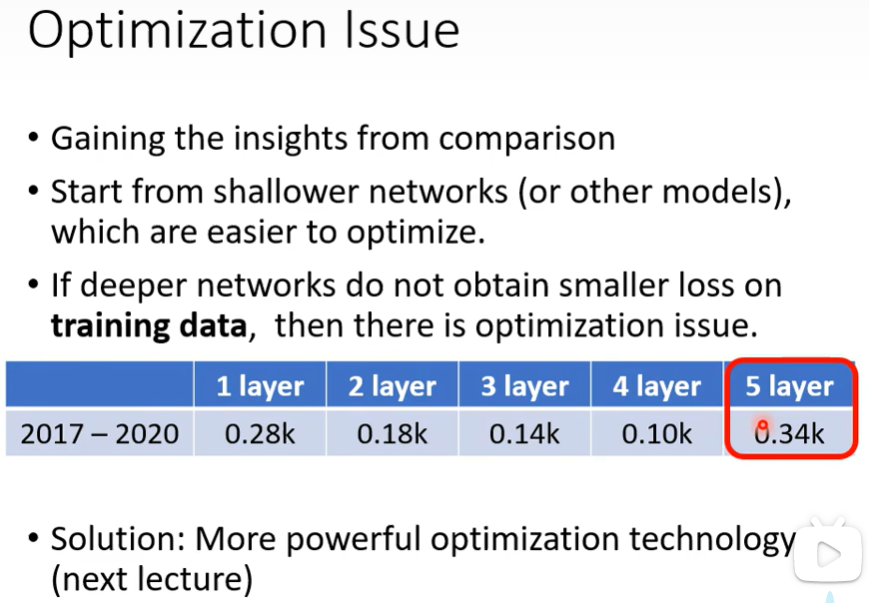


**对于model bias**

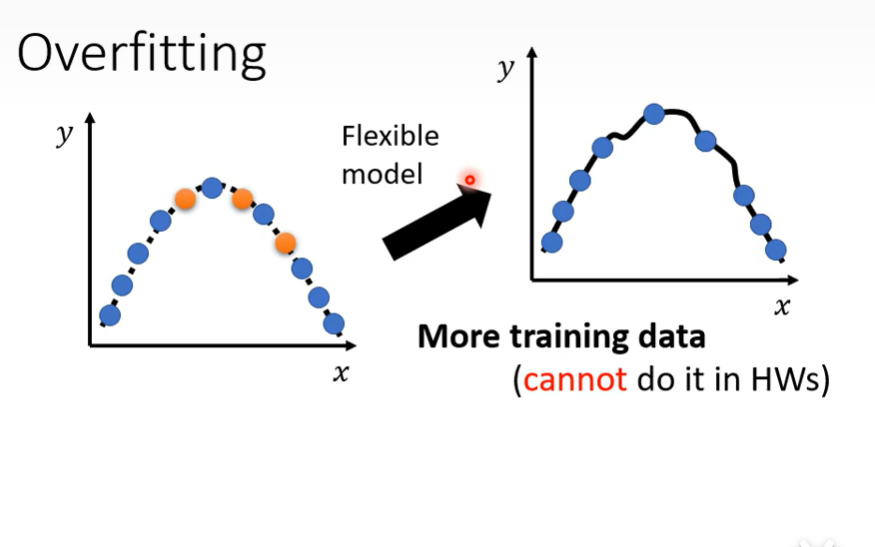


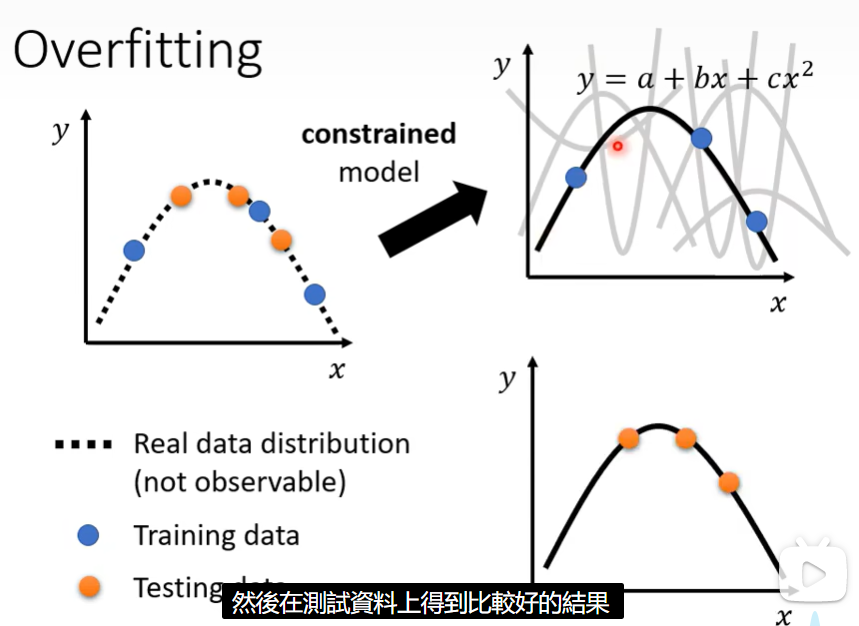
**对于optimization issue**

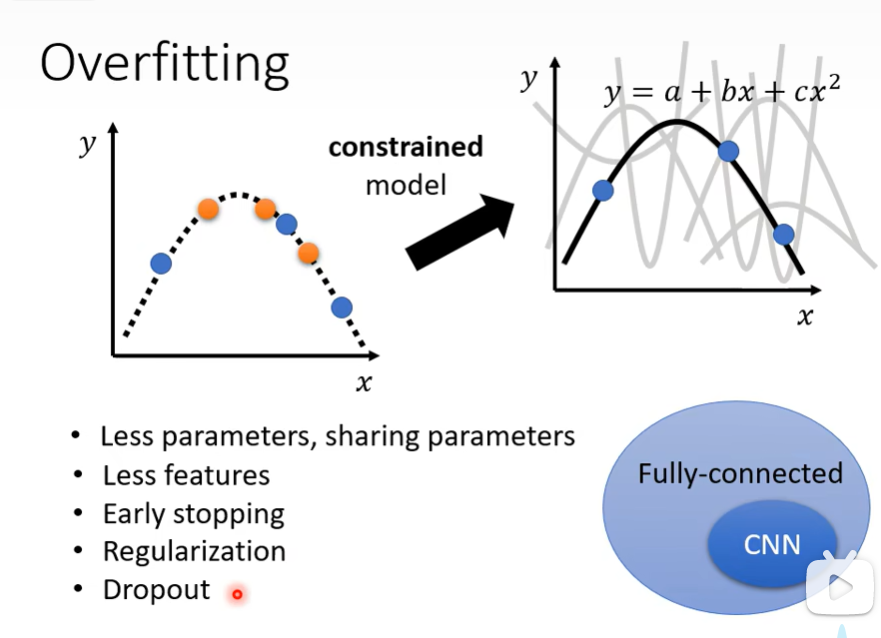


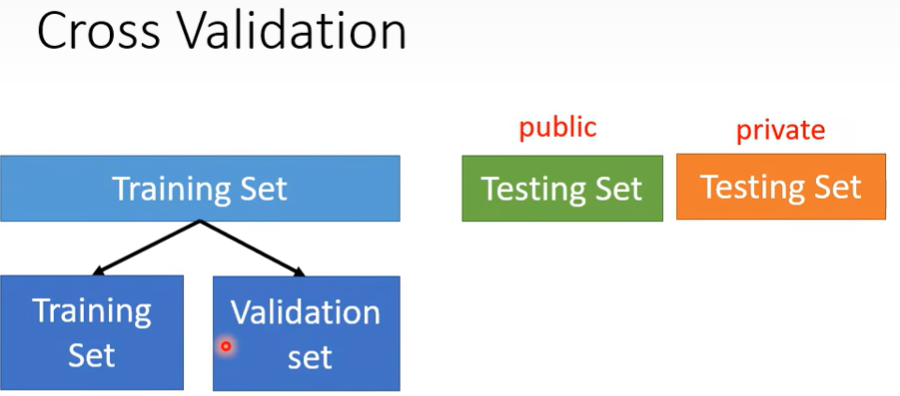


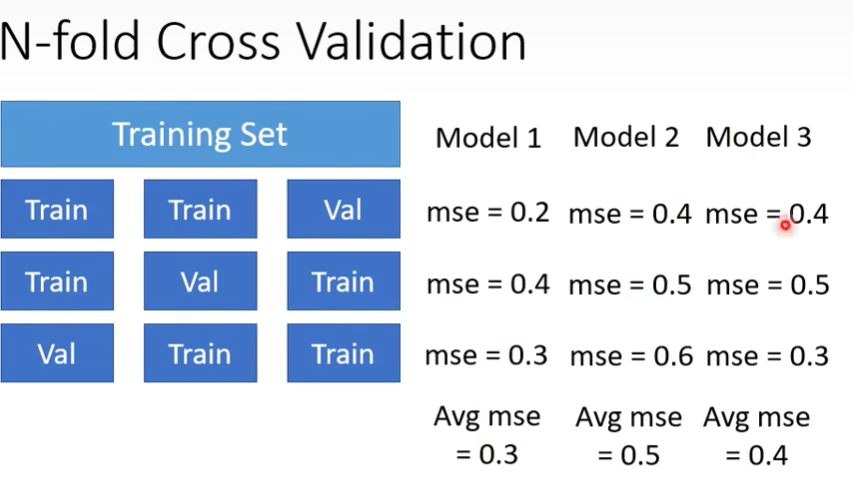
**对于overfitting**



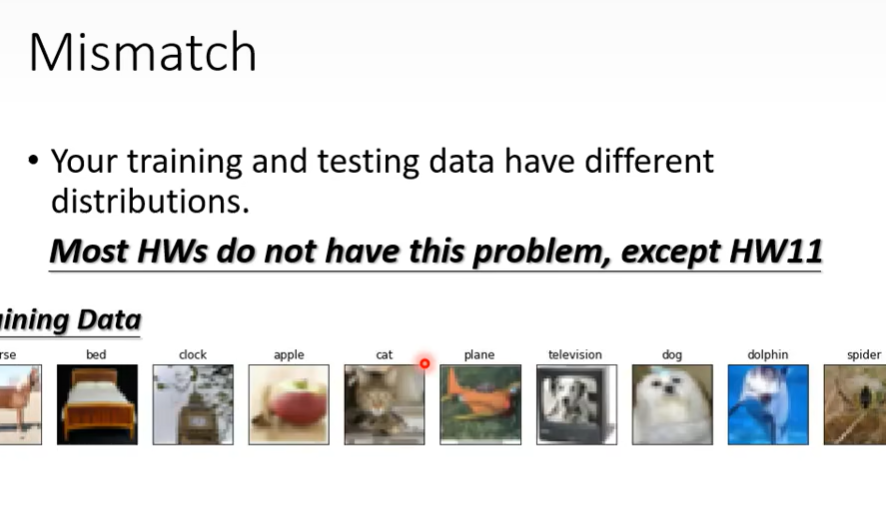
****



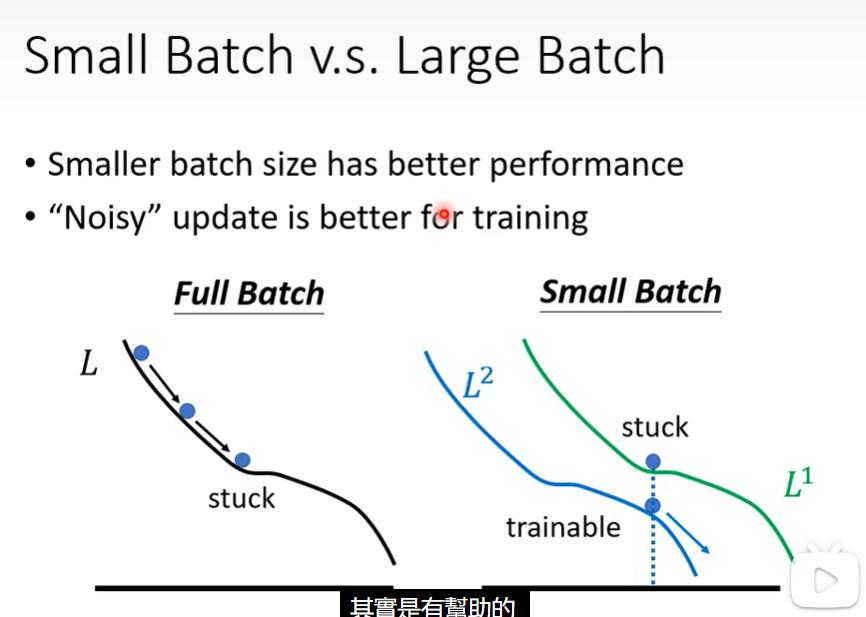


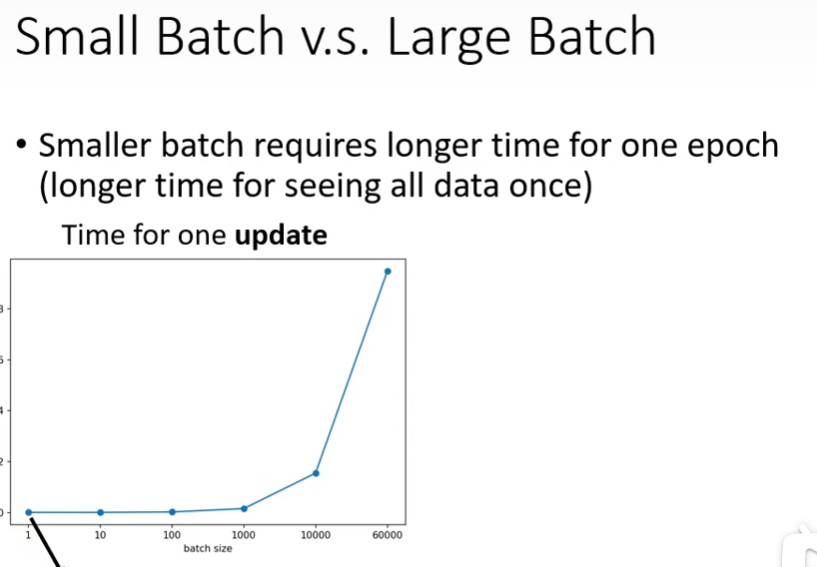


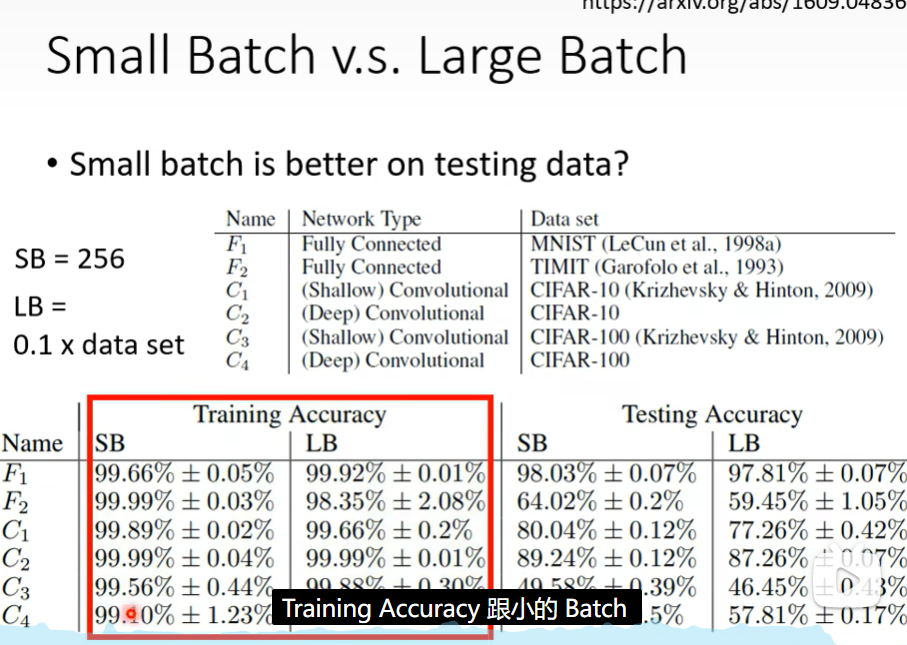
**关于Mismatch**

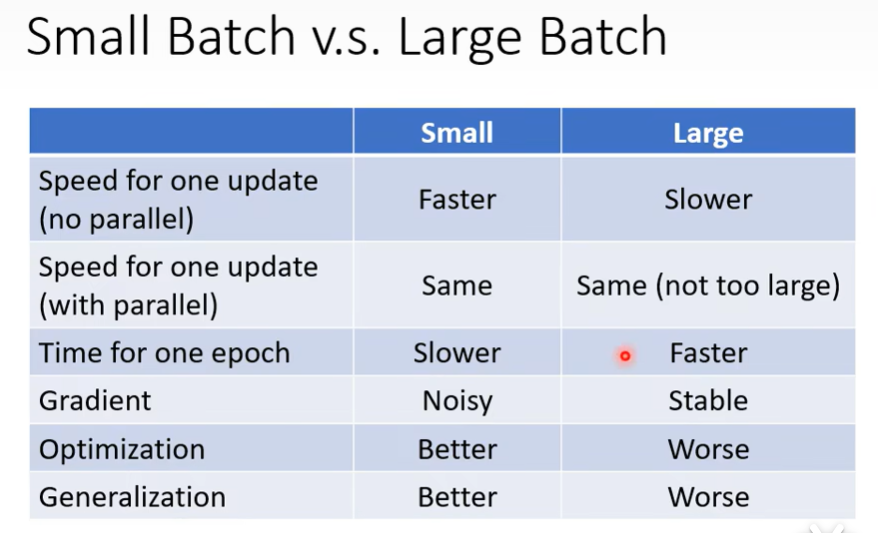


**Small Batch v.s. Large Batch**

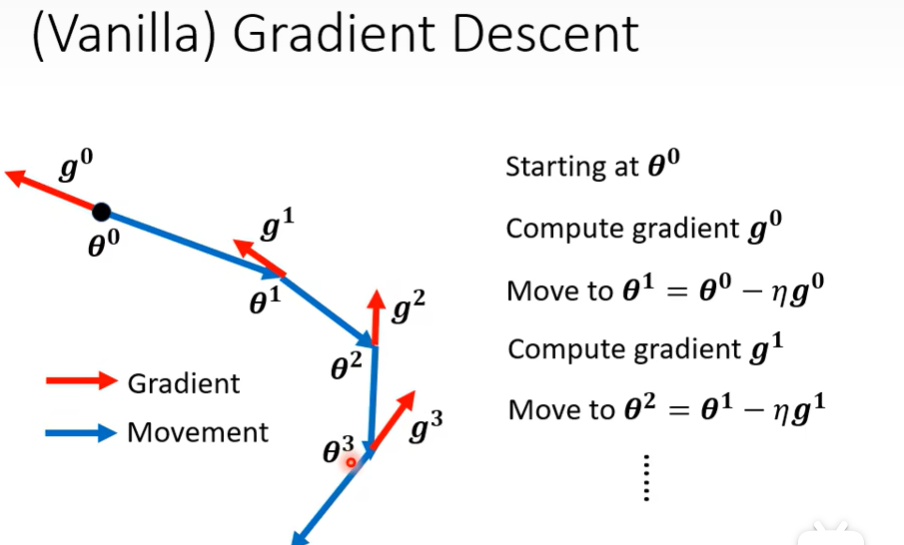


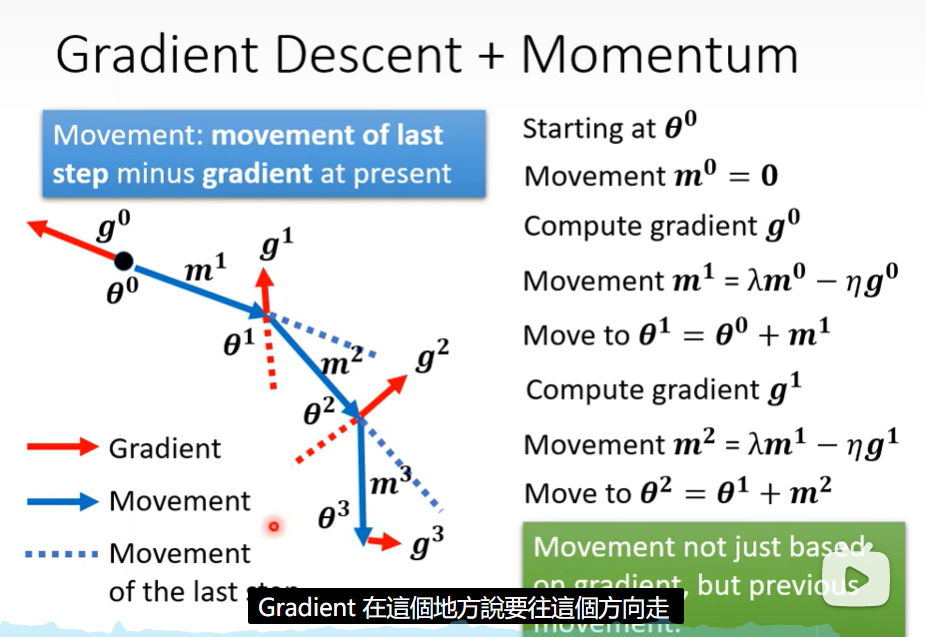




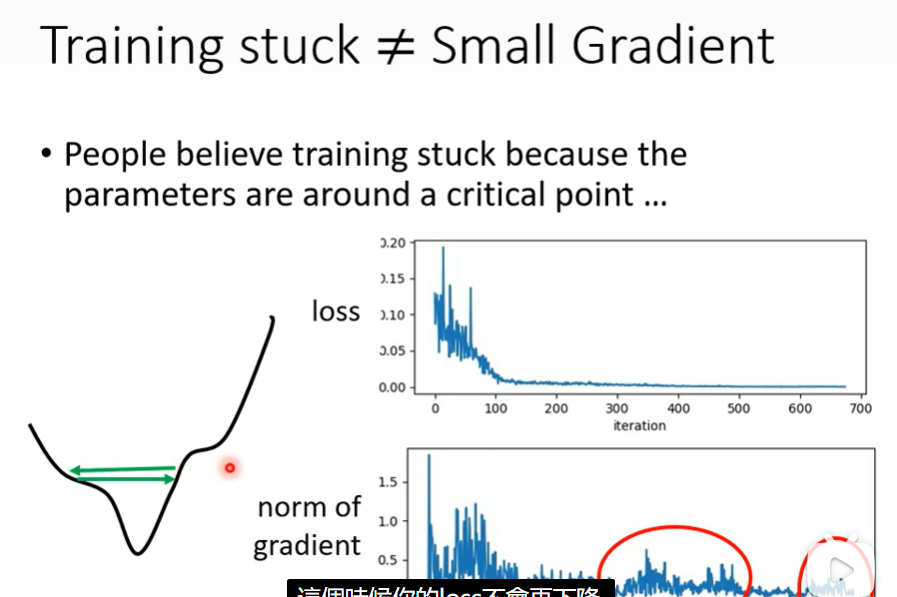


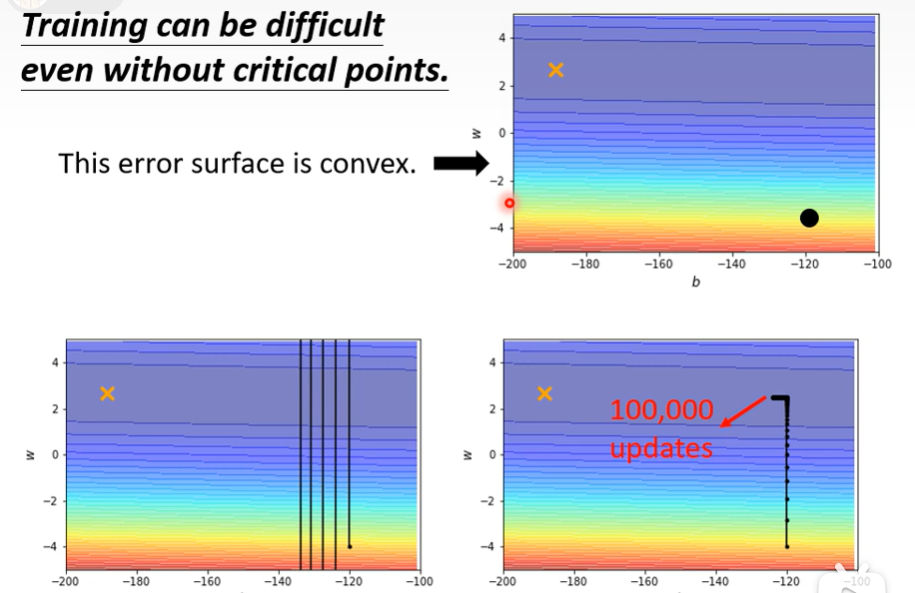
**Momentum**

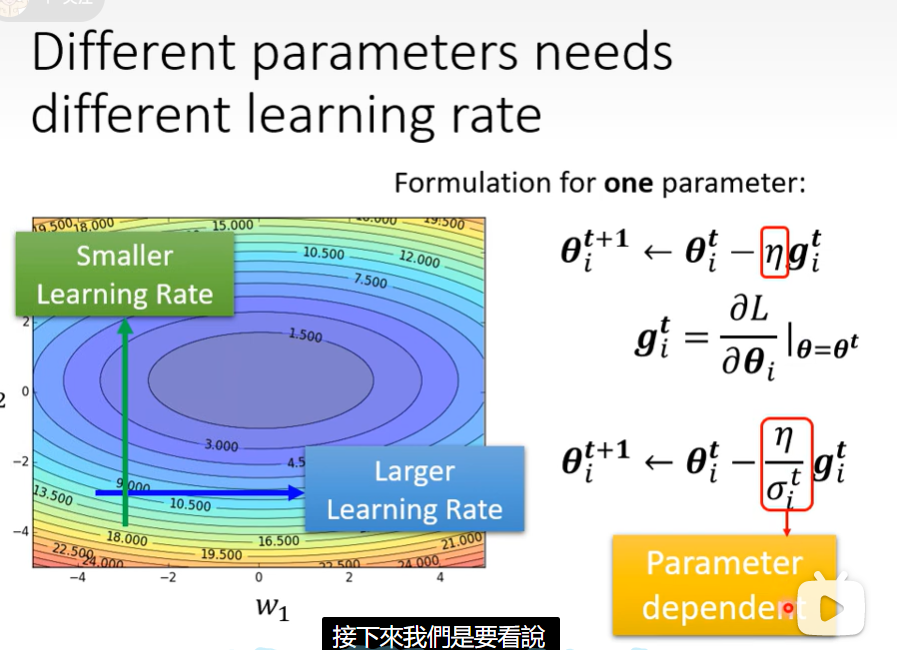




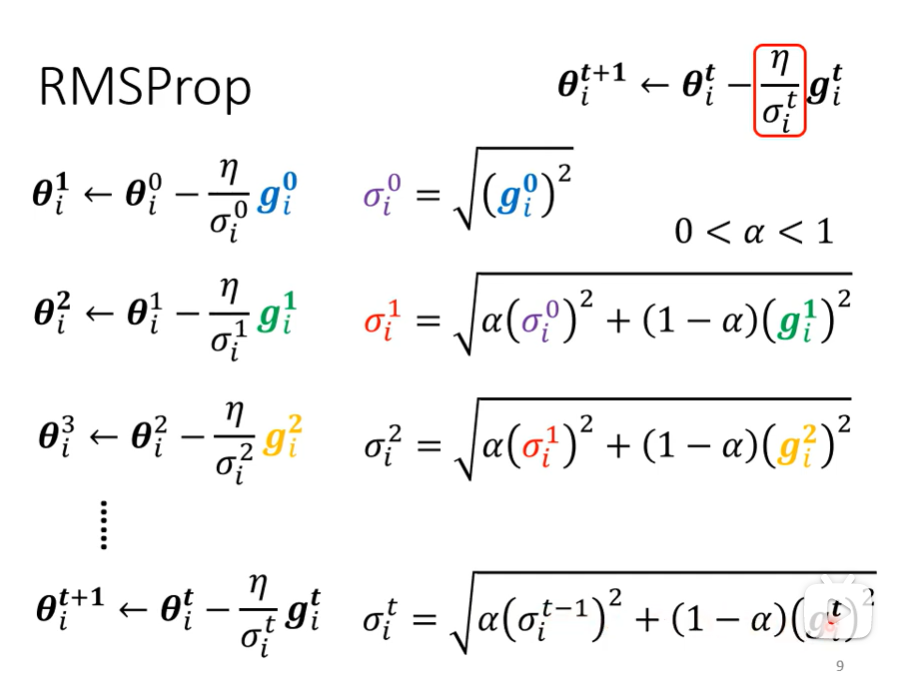
**Adaptive Learning Rate**

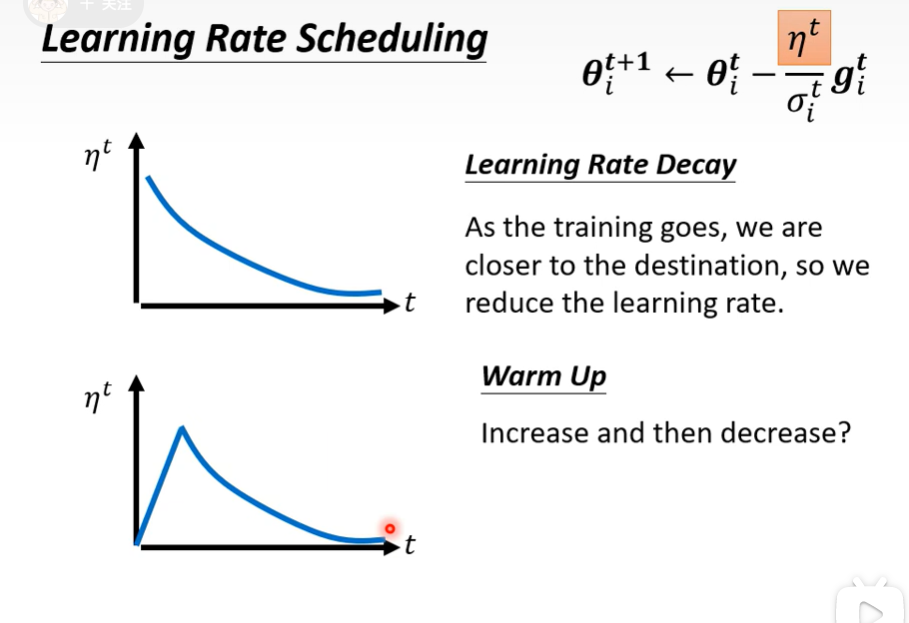
****



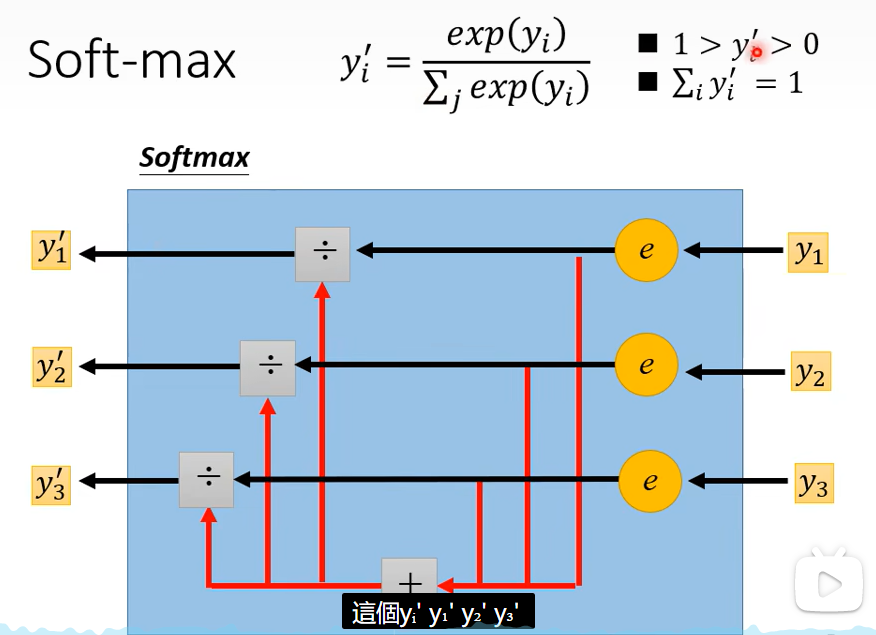


**RMSProp**

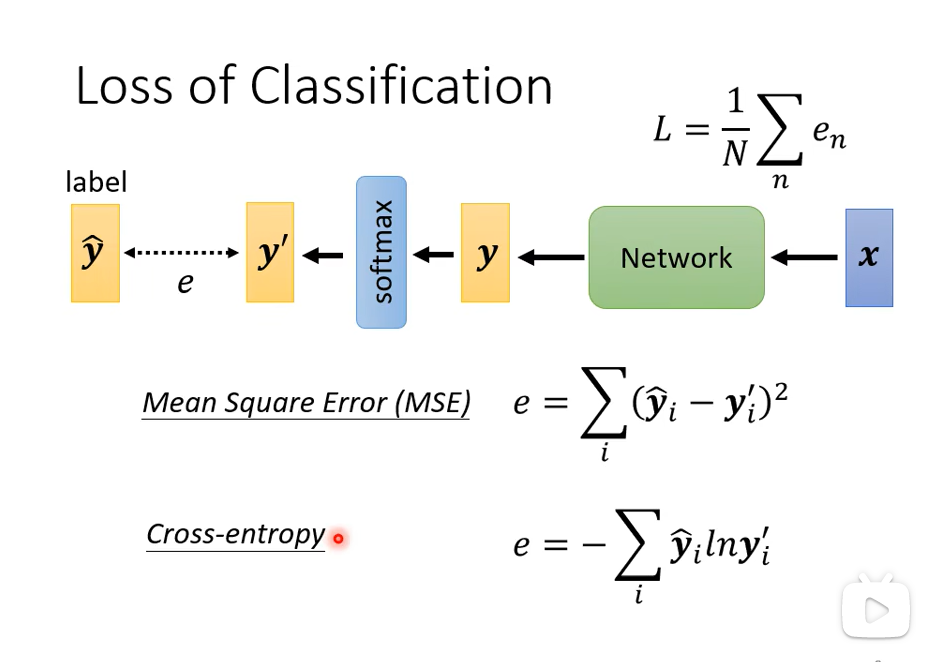


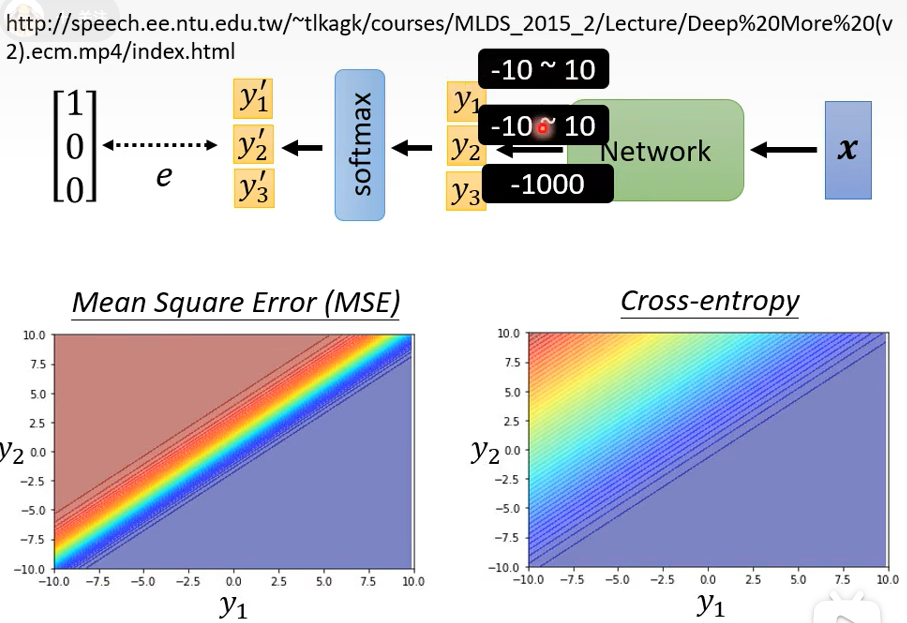


**Soft-max**

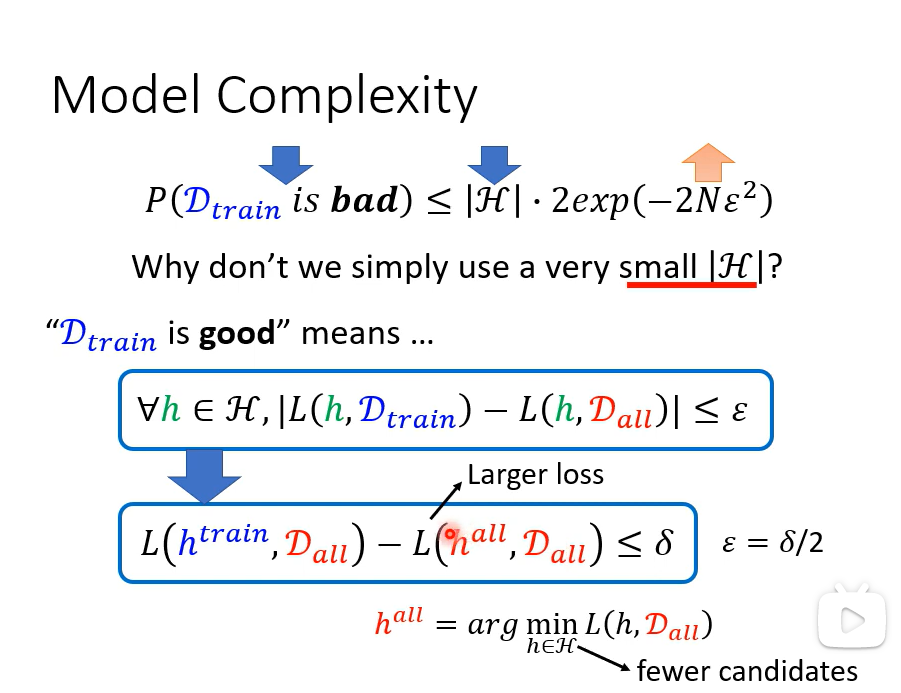


**MSE与Cross-entropy**

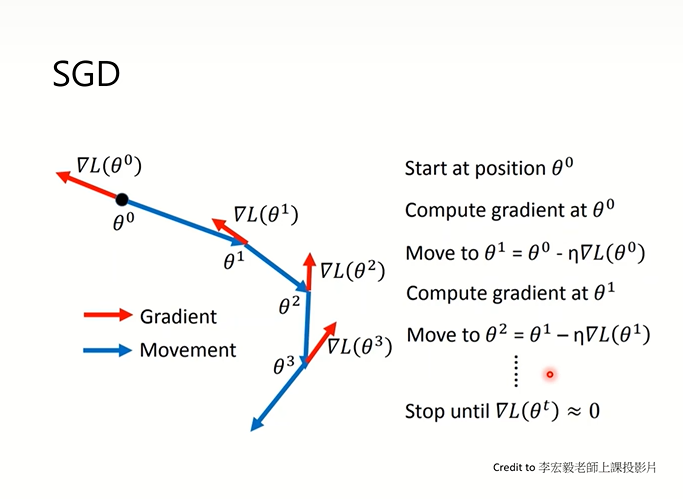




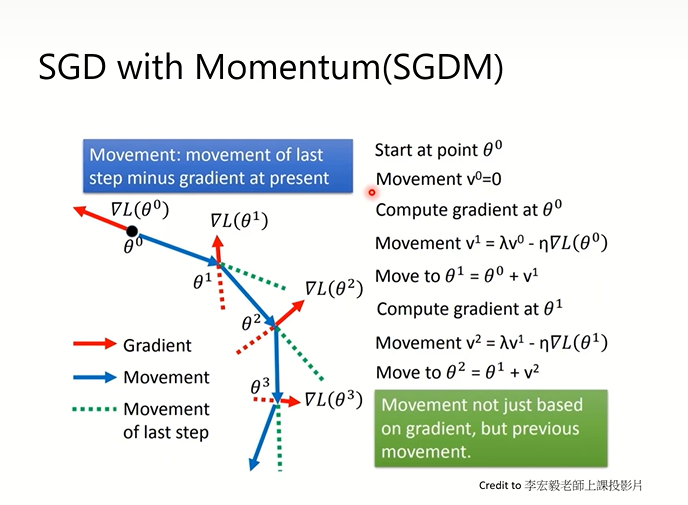
**Model Complexity**



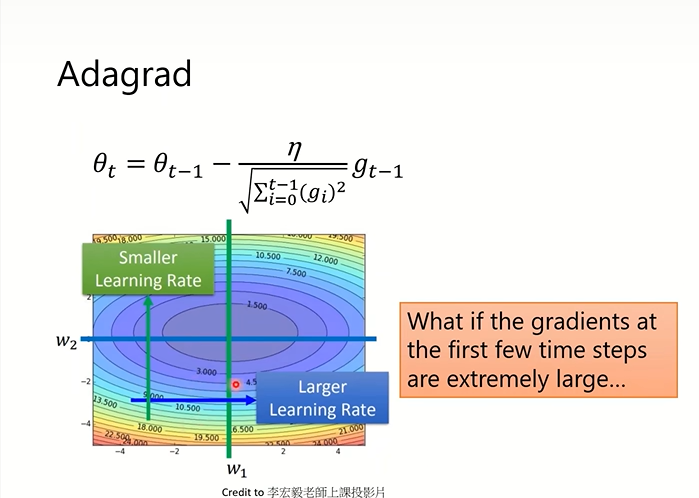
**SGD**



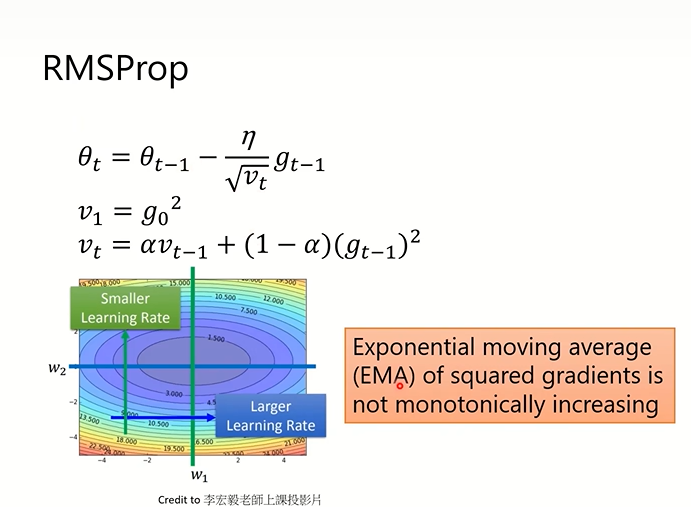
**SGDM**



**Adagrad**



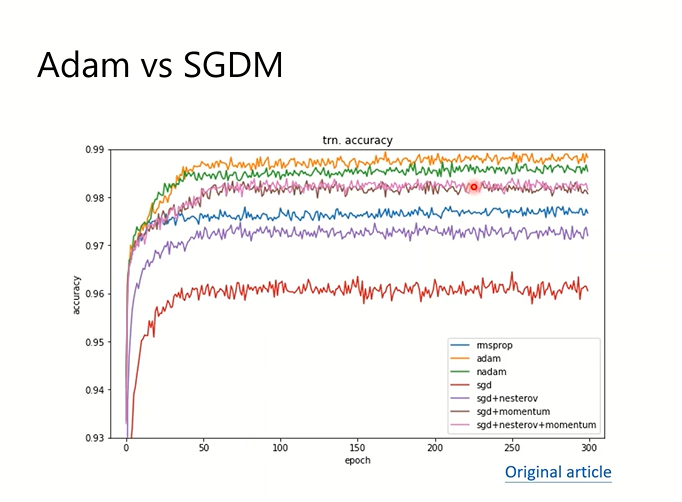
**RMSProp**

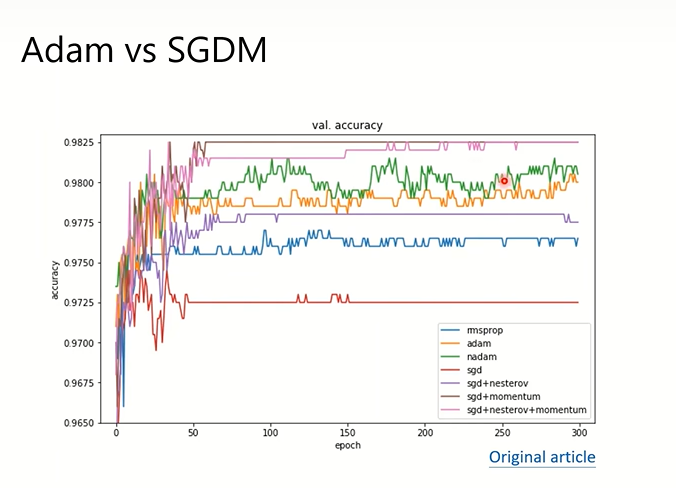


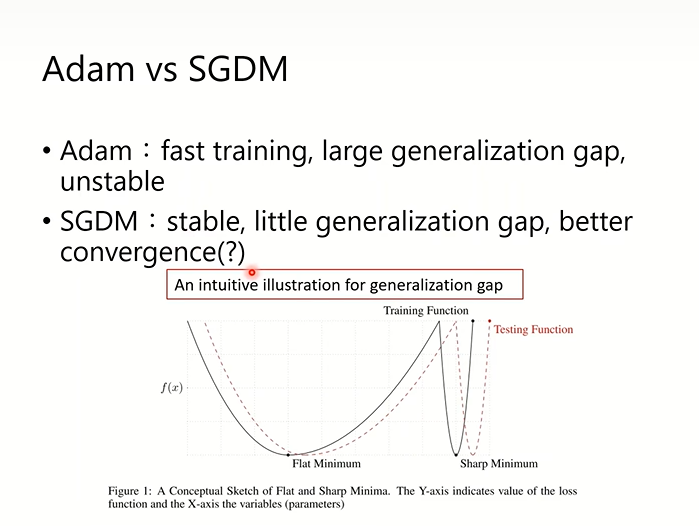
**Adam**



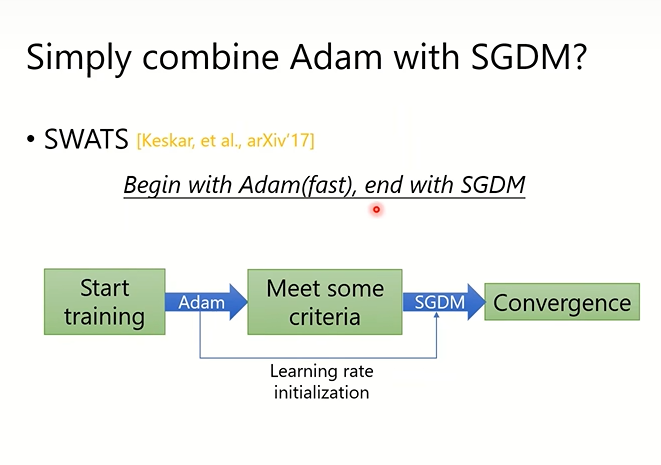
**Adam vs SGDM**



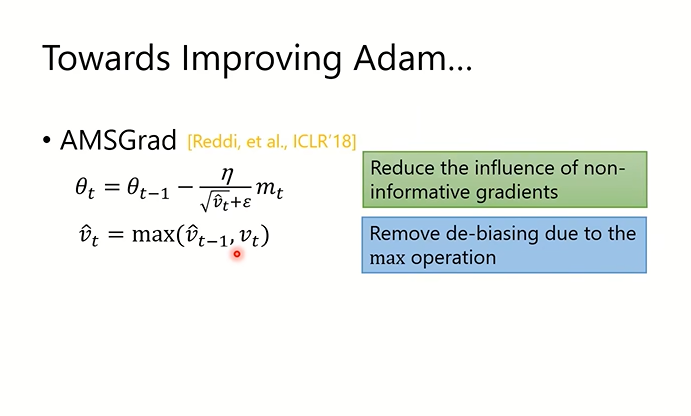




**SWATS**



**AMSGrad**



**AdaBound**

