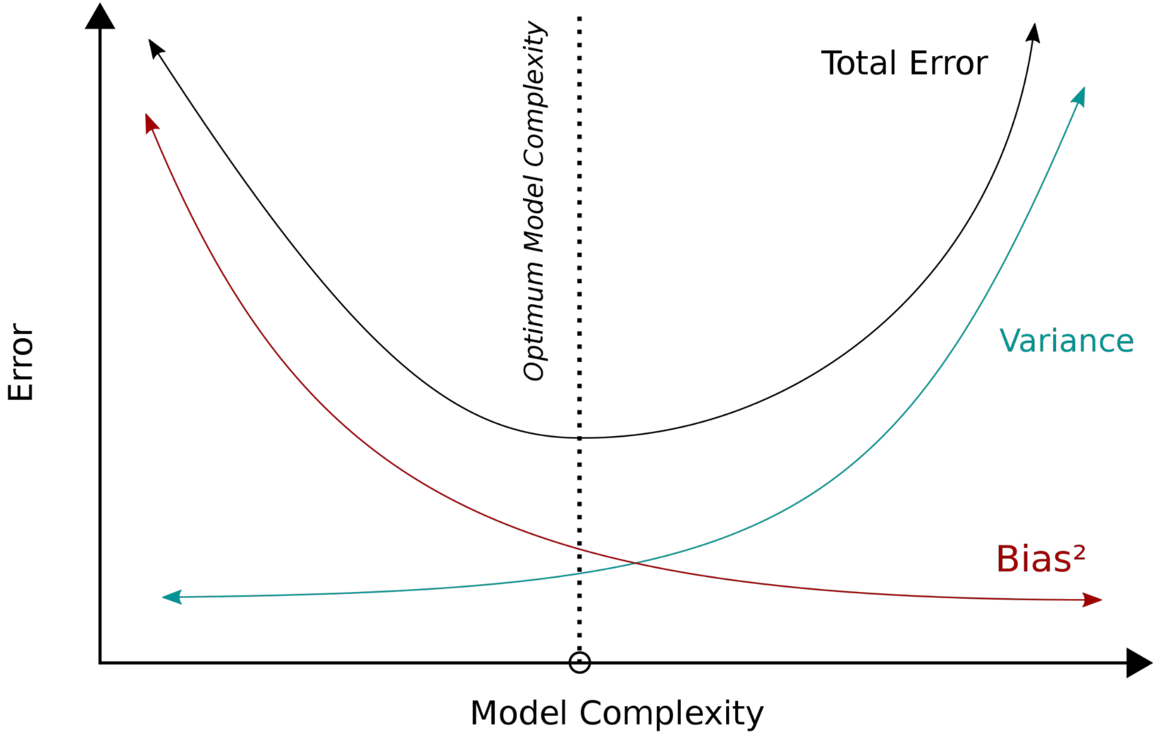
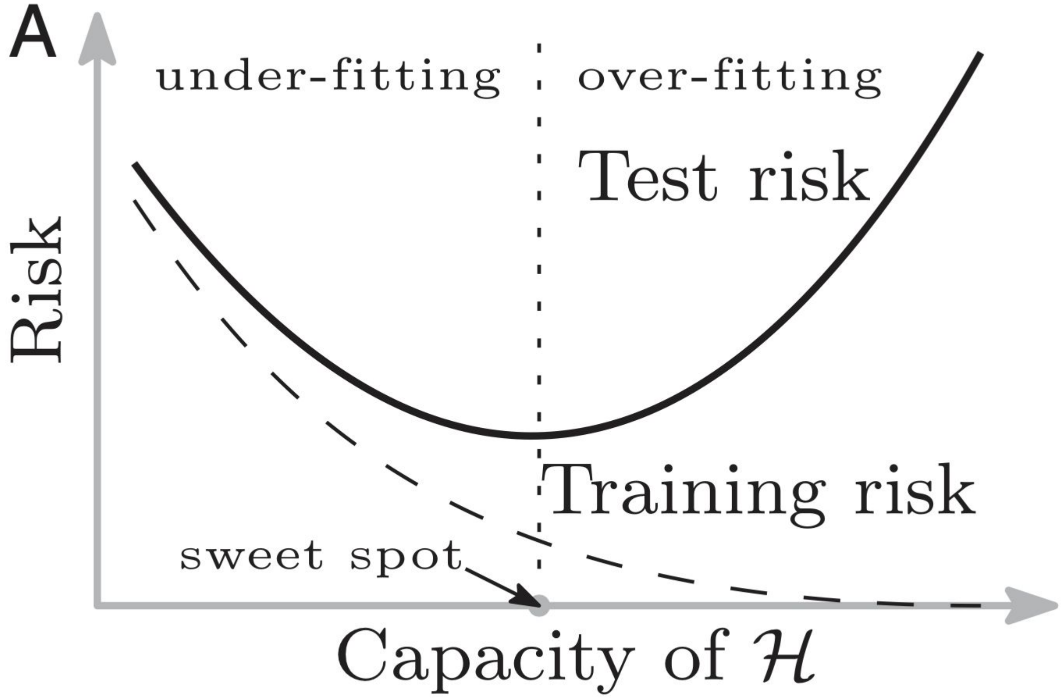
Modern ML / DL vs Classical Bias-Variance Tradeoff

Classical Bias vs Variance Tradeoff



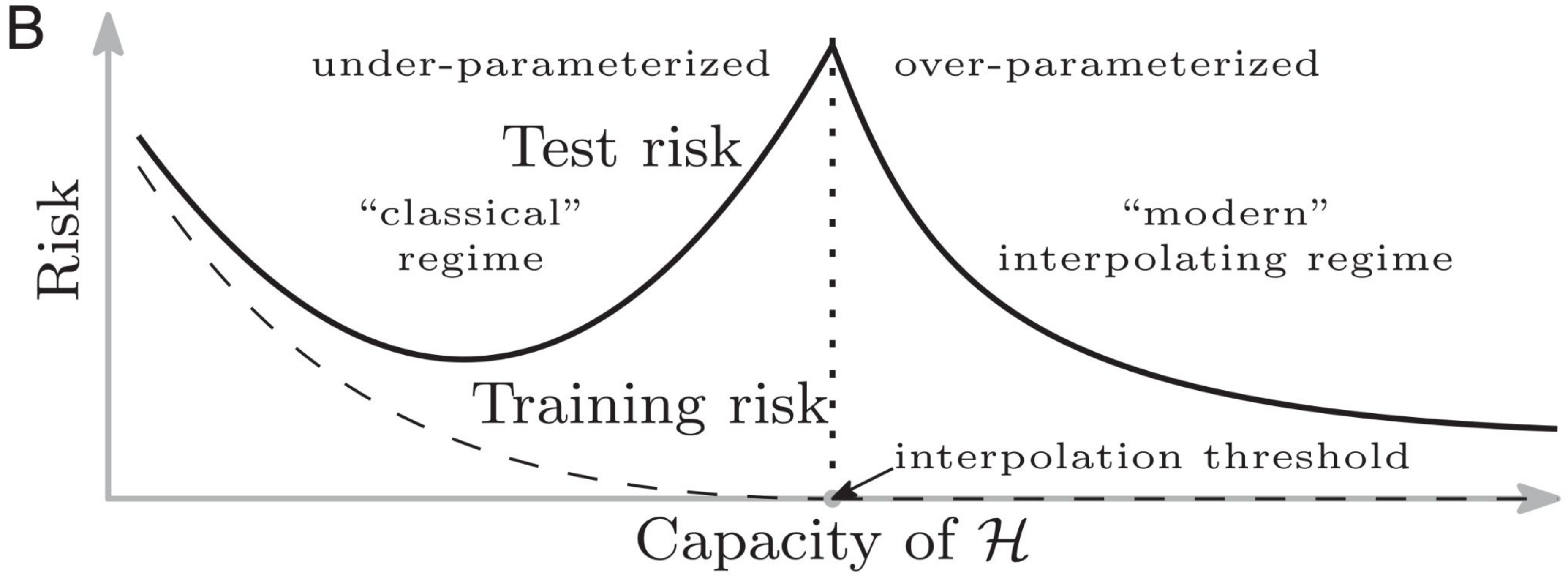
https://en.wikipedia.org/wiki/Bias%E2%80%93variance_tradeoff

Classical: Under-fitting vs Over-fitting



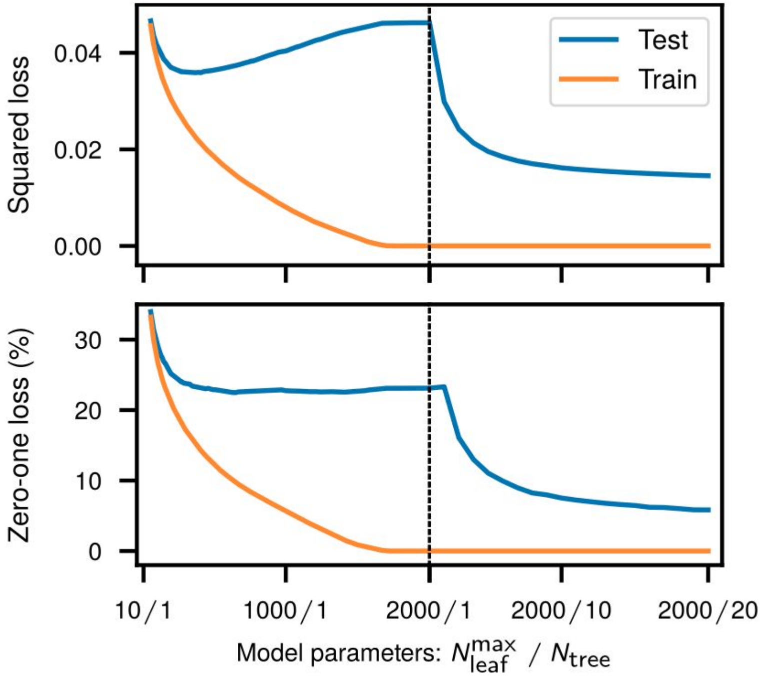
Belkin, M., Hsu, D., Ma, S., & Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, *116*(32), 15849-15854.

Modern: Under-parameterized vs Over-parameterized



Belkin, M., Hsu, D., Ma, S., & Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, *116*(32), 15849-15854.

Double Descent - Random Forest

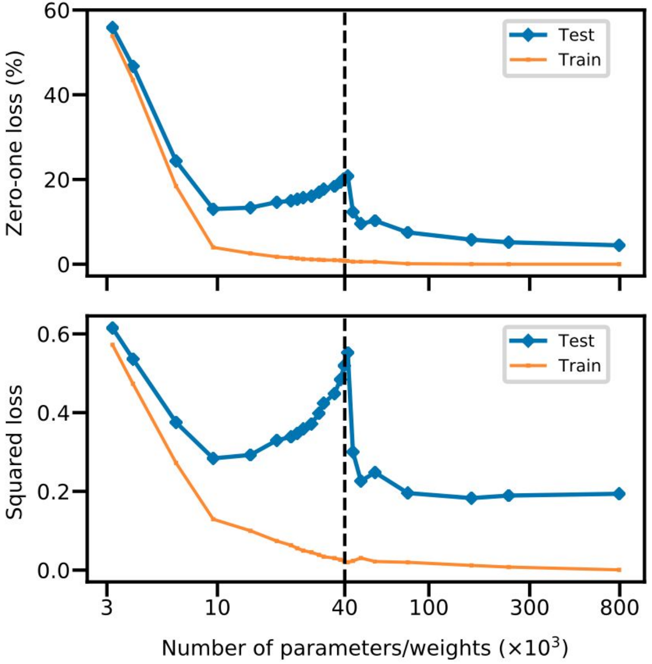


The double-descent risk curve is observed for random forests with increasing model complexity trained on a subset of MNIST (n = 104 , 10 classes).

Its complexity is controlled by the number of trees (N *tree*) and the maximum number of leaves allowed for each tree (*N max leaf*)

Belkin, M., Hsu, D., Ma, S., & Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, *116*(32), 15849-15854.

Double Descent - Fully Connected Neural Network



Double-descent risk curve for a fully connected neural network on MNIST. Shown are training and test risks of a network with a single layer of H hidden units, learned on a subset of MNIST (n = 4 x 10^3 , d = 784, K=10 classes).

The number of parameters is (d+1)·H+(H+1)·K.

The interpolation threshold (black dashed line) is observed at n · K.

Belkin, M., Hsu, D., Ma, S., & Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, *116*(32), 15849-15854.