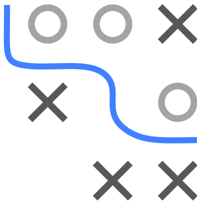
WHY/WHEN DOES BAGGING HELP?



Assume we use quadratic loss and measure instability of the ensemble with $\Delta\left(f^{[M]}(\mathbf{x})\right) = \frac{1}{M}\sum_m^M \left(b^{[m]} - f^{[M]}(\mathbf{x})\right)^2$:

$$
\begin{aligned}
\Delta\left(f^{[M]}(\mathbf{x})\right) &= \frac{1}{M}\sum_m^M \left(b^{[m]} - f^{[M]}(\mathbf{x})\right)^2 \\
&= \frac{1}{M}\sum_m^M \left(\left(b^{[m]} - y\right) + \left(y - f^{[M]}(\mathbf{x})\right)\right)^2 \\
&= \frac{1}{M}\sum_m^M L(y, b^{[m]}) + L(y, f^{[M]}(\mathbf{x})) \underbrace{-2\left(y - \frac{1}{M}\sum_{m=1}^M b^{[m]}\right)\left(y - f^{[M]}(\mathbf{x})\right)}_{-2L\left(y, f^{[M]}(\mathbf{x})\right)}
\end{aligned}
$$

So, if we take the expected value over the data's distribution:

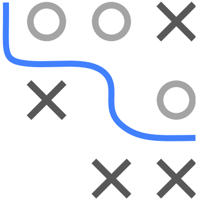
$$
\mathbb{E}_{xy}\left[L\left(y, f^{[M]}(\mathbf{x})\right)\right] = \frac{1}{M}\sum_m^M \mathbb{E}_{xy}\left[L\left(y, b^{[m]}\right)\right] - \mathbb{E}_{xy}\left[\Delta\left(f^{[M]}(\mathbf{x})\right)\right]
$$

$\Rightarrow$ The expected loss of the ensemble is lower than the average loss of the single base learner by the amount of instability in the ensemble's base learners.

The more accurate and diverse the base learners, the better.

IMPROVING BAGGING

How to make $\mathbb{E}_{xy}\left[\Delta\left(f^{[M]}(\mathbf{x})\right)\right]$ as large as possible?



$$\mathbb{E}_{xy}\left[L\left(y, f^{[M]}(\mathbf{x})\right)\right] = \frac{1}{M}\sum_{m}^{M}\mathbb{E}_{xy}\left[L\left(y, b^{[m]}\right)\right] - \mathbb{E}_{xy}\left[\Delta\left(f^{[M]}(\mathbf{x})\right)\right]$$

Assume $\mathbb{E}_{xy}\left[b^{[m]}\right] = 0$ for simplicity, $\text{Var}_{xy}\left[b^{[m]}\right] = \mathbb{E}_{xy}\left[(b^{[m]})^2\right] = \sigma^2$,
$\text{Corr}_{xy}\left[b^{[m]}, b^{[m]}m'\right] = \rho$ for all m, m'.

$$\implies \text{Var}_{xy}\left[f^{[M]}(\mathbf{x})\right] = \frac{1}{M}\sigma^2 + \frac{M-1}{M}\rho\sigma^2 \qquad \left(\ldots = \mathbb{E}_{xy}\left[(f^{[M]}(\mathbf{x}))^2\right]\right)$$

$$\begin{aligned}
\mathbb{E}_{xy}\left[\Delta\left(f^{[M]}(\mathbf{x})\right)\right] &= \frac{1}{M}\sum_{m}^{M}\mathbb{E}_{xy}\left[\left(b^{[m]} - f^{[M]}(\mathbf{x})\right)^2\right] \\
&= \frac{1}{M}\left(M\mathbb{E}_{xy}\left[(b^{[m]})^2\right] + M\mathbb{E}_{xy}\left[(f^{[M]}(\mathbf{x}))^2\right] - 2M\mathbb{E}_{xy}\left[b^{[m]}f^{[M]}(\mathbf{x})\right]\right) \\
&= \sigma^2 + \mathbb{E}_{xy}\left[(f^{[M]}(\mathbf{x}))^2\right] - 2\frac{1}{M}\sum_{m'}^{M}\underbrace{\mathbb{E}_{xy}\left[b^{[m]}b^{[m]}m'\right]}_{=\text{Cov}_{xy}\left[b^{[m]}, b^{[m]}m'\right] + \mathbb{E}_{xy}\left[b^{[m]}\right]\mathbb{E}_{xy}\left[b^{[m]}m'\right]} \\
&= \sigma^2 + \left(\frac{1}{M}\sigma^2 + \frac{M-1}{M}\rho\sigma^2\right) - 2\left(\frac{M-1}{M}\rho\sigma^2 + \frac{1}{M}\sigma^2 + 0\cdot 0\right) \\
&= \frac{M-1}{M}\sigma^2(1-\rho)
\end{aligned}$$