Some small proofs

For those who are interested we give some small proofs to go with the lecture slides. In these proofs we use the properties of expectation listed at the end.

Irreducible error

First we show that if we want to minimize expected squared prediction error $\mathbb{E}_{P(Y \mid x)}\left[(Y-c)^2\right]$, then we should choose $c = \mathbb{E}[Y \mid x]$. We will suppress dependence on x to simplify the notation. First, we expand the square, en push the expectation operator inwards:

$$\begin{aligned}\mathbb{E}\left[(Y-c)^2\right] &= \mathbb{E}\left[Y^2 - 2cY + c^2\right] \\ &= \mathbb{E}\left[Y^2\right] - 2c\mathbb{E}[Y] + c^2\end{aligned}$$

Note that c is a variable in the normal mathematical sense, but it is not a random variable. The only random variable here is Y (and any function of Y of course, in this case Y^2). So when it comes to taking expectations, c can be treated as a constant.

Next, we take the derivative of this expression with respect to c:

$$\frac{d\,\mathbb{E}\left[Y^2\right] - 2c\mathbb{E}[Y] + c^2}{d\,c} = -2\mathbb{E}[Y] + 2c$$

Equating the derivative to zero gives

$$\begin{aligned}-2\mathbb{E}[Y] + 2c &= 0 \\ c &= \mathbb{E}[Y]\end{aligned}$$

Note that the second derivative is positive everywhere, so we are dealing with a global minimum. Since this argument works for arbitrary values of x, we have established that the population regression function

$$f(x) = \mathbb{E}\left[Y \mid x\right]$$

yields the best possible prediction for Y. Its expected squared error is

$$\mathbb{E}\left[(Y - \mathbb{E}[Y \mid x])^2\right]$$

which is the variance of Y for $X = x$.

Decomposition into Irreducible and Reducible error

We suppress dependence on x to simplify the notation. We stated that for fixed x and $\hat{f}$:

$$\mathbb{E}[(Y - \hat{f})^2] = (f - \hat{f})^2 + \mathbb{E}[(Y - f)^2]$$

Proof:

$$\begin{aligned}
\mathbb{E}[(Y - \hat{f})^2] &= \mathbb{E}[((Y - f) + (f - \hat{f}))^2] && \text{(Add and subtract } f\text{)} \\
&= \mathbb{E}[(Y - f)^2 + 2(Y - f)(f - \hat{f}) + (f - \hat{f})^2] && \text{(Expand the square)} \\
&= \mathbb{E}[(Y - f)^2] + 2(\mathbb{E}[Y] - f)(f - \hat{f}) + (f - \hat{f})^2 && \text{(Push expectation inside)} \\
&= \mathbb{E}[(Y - f)^2] + (f - \hat{f})^2
\end{aligned}$$

In the last step, $2(\mathbb{E}[Y] - f)(f - \hat{f})$ drops out, because $f = \mathbb{E}[Y]$, and therefore $(\mathbb{E}[Y] - f)$ evaluates to zero.

Error Reduction by Bagging

The average error made by the models acting individually is:

$$E_{\text{AV}} = \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}\left[e_m(x)^2\right]$$

Similarly, the expected error from the committee is given by

$$\begin{aligned}
E_{\text{BAG}} &= \mathbb{E}\left[\left(\frac{1}{M}\sum_{m=1}^{M} \hat{f}_m(x) - f(x)\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{1}{M}\sum_{m=1}^{M} e_m(x)\right)^2\right]
\end{aligned}$$

If we assume that the errors have zero mean and are uncorrelated, so that

$$\mathbb{E}\left[e_m(x)e_n(x)\right] = 0, \text{ for all } m \neq n$$

then we obtain

$$E_{\text{BAG}} = \frac{1}{M} E_{\text{AV}}$$

Proof:

$$
\begin{aligned}
E_{\text{BAG}} &= \mathbb{E}\left[\left(\frac{1}{M}\sum_{m=1}^{M} e_m(x)\right)^2\right] \\
&= \frac{1}{M^2}\,\mathbb{E}\left[\left(\sum_{m=1}^{M} e_m(x)\right)^2\right] \\
&= \frac{1}{M^2}\,\mathbb{E}\left[\sum_{m=1}^{M} e_m(x)^2 + \sum_{m\neq n} e_m(x)e_n(x)\right] \\
&= \frac{1}{M^2}\,\left(\sum_{m=1}^{M} \mathbb{E}\left[e_m(x)^2\right] + \sum_{m\neq n} \mathbb{E}\left[e_m(x)e_n(x)\right]\right) \\
&= \frac{1}{M^2}\,\sum_{m=1}^{M} \mathbb{E}\left[e_m(x)^2\right] = \frac{1}{M}E_{\text{AV}}
\end{aligned}
$$

In the last step we used the assumption that

$$\mathbb{E}\left[e_m(x)e_n(x)\right] = 0, \text{ for all } m \neq n.$$

Some Useful Properties of Expectation and Variance

1. $\mathbb{E}[c] = c$ for constant c. “The expected value of a constant is the constant itself”.
2. $\mathbb{E}[cX] = c\mathbb{E}[X]$.
3. $\mathbb{E}[X \pm Y] = \mathbb{E}[X] \pm \mathbb{E}[Y]$.
4. $\mathbb{V}[c] = 0$ for constant c. “The variance of a constant is zero”.
5. $\mathbb{V}[cX] = c^2\,\mathbb{V}[X]$. “The variance of a constant times a random variable is equal to the square of the constant times the variance of the random variable”.
6. $\mathbb{V}[X \pm Y] = \mathbb{V}[X] + \mathbb{V}[Y]$ if X and Y are independent.

More generally, let $Z = c_0 + \sum_{i=1}^{n} c_i X_i$. Then

1. $\mathbb{E}[Z] = \mathbb{E}\left[c_0 + \sum_{i=1}^{n} c_i X_i\right] = c_0 + \sum_{i=1}^{n} c_i \mathbb{E}[X_i]$
2. $\mathbb{V}[Z] = \mathbb{V}\left[c_0 + \sum_{i=1}^{n} c_i X_i\right] = \sum_{i=1}^{n} c_i^2\,\mathbb{V}[X_i]$, provided that the X_i are mutually independent.