part2 - a

$$
\begin{aligned}
D &= E(\frac{1}{n}\|f - \hat{f}\|^2) \\
&= E(\frac{1}{n}\|f - E(\hat{f}) - \hat{f} + E(\hat{f})\|^2) \\
&= E(\frac{1}{n}\|f - E(\hat{f}) - \hat{f} + E(\hat{f})\|^2) \\
&= \frac{1}{n}[E(\|f - E(\hat{f})\|^2) + E(\|\hat{f} - E(\hat{f})\|^2) + 2E(E(\|f - E(\hat{f})\|)E(\|\hat{f} - E(\hat{f})\|)]
\end{aligned}
$$

Note that $E(E(\|\hat{f} - E(\hat{f})\|)) = 0$, two terms are independent to each other, therefore the cross term $2E(E(\|f - E(\hat{f})\|)E(\|\hat{f} - E(\hat{f})\|) = 0$.

$$
D = \frac{1}{n}[E(\|f - E(\hat{f})\|^2) + E(\|\hat{f} - E(\hat{f})\|^2)]
$$

Separately write two terms.

Since $\hat{f} = Wy$ and $y = f + \epsilon$, where y, f, ϵ are vectors, $E(\hat{f}) = E(Wy) = WE(f + \epsilon) = Wf$

$$
\begin{aligned}
E(\|f - E(\hat{f})\|^2) &= E(\|f - Wf\|^2) \\
&= E(\|(W - I)f\|^2)
\end{aligned}
$$

Since W depends only on bandwidth, and f is parameter, this term can be written as $\|(W - I)f\|^2$

$E(\|\hat{f} - E(\hat{f})\|^2)$ is the variance of $\|\hat{f}\|$, where $\hat{f} = Wy$.

$$
\begin{aligned}
E(\|\hat{f} - E(\hat{f})\|^2) &= Var(\|W\|y) \\
&= \|W^2\|Var(y) \\
&= trace(W^T W)\sigma^2
\end{aligned}
$$

Combine these two term, we proved:

$$
D = \frac{1}{n}(\|(W - I)f\|^2 + trace(W^T W)\sigma^2)
$$