5.17

$$E = \frac{1}{2} \int \int \{y(\mathbf{x}, \mathbf{w}) - t\}^2 \, p(\mathbf{x}, t) \, d\mathbf{x} \, dt$$

$$\implies \frac{\partial E}{\partial w_s} = \int \int \{y(\mathbf{x}, \mathbf{w}) - t\} \left(\frac{\partial y(\mathbf{x}, \mathbf{w})}{\partial w_s} \right) p(\mathbf{x}, t) \, d\mathbf{x} \, dt$$

$$\implies \frac{\partial^2 E}{\partial w_r \partial w_s} = \int \int \left(\left(\frac{\partial y(\mathbf{x}, \mathbf{w})}{\partial w_r} \right) \left(\frac{\partial y(\mathbf{x}, \mathbf{w})}{\partial w_s} \right) \right.$$
$$\left. + \{y(\mathbf{x}, \mathbf{w}) - t\} \left(\frac{\partial^2 y(\mathbf{x}, \mathbf{w})}{\partial w_r \partial w_s} \right) \right) p(\mathbf{x}, t) \, d\mathbf{x} \, dt$$

$$= \int \int \left(\frac{\partial y(\mathbf{x}, \mathbf{w})}{\partial w_r} \right) \left(\frac{\partial y(\mathbf{x}, \mathbf{w})}{\partial w_s} \right) p(\mathbf{x}, t) \, d\mathbf{x} \, dt$$
$$+ \int \int \{y(\mathbf{x}, \mathbf{w}) - t\} \left(\frac{\partial^2 y(\mathbf{x}, \mathbf{w})}{\partial w_r \partial w_s} \right) p(\mathbf{x}, t) \, d\mathbf{x} \, dt$$

$$= \int \left(\frac{\partial y(\mathbf{x}, \mathbf{w})}{\partial w_r} \right) \left(\frac{\partial y(\mathbf{x}, \mathbf{w})}{\partial w_s} \right) p(\mathbf{x}) \, d\mathbf{x}$$
$$+ \int \int y(\mathbf{x}, \mathbf{w}) \left(\frac{\partial^2 y(\mathbf{x}, \mathbf{w})}{\partial w_r \partial w_s} \right) p(\mathbf{x}, t) \, d\mathbf{x} \, dt$$
$$- \int \int t \left(\frac{\partial^2 y(\mathbf{x}, \mathbf{w})}{\partial w_r \partial w_s} \right) p(\mathbf{x}, t) \, d\mathbf{x} \, dt$$

Simplifying the second term by integrating out t:

$$\int \int y(\mathbf{x}, \mathbf{w}) \left(\frac{\partial^2 y(\mathbf{x}, \mathbf{w})}{\partial w_r \partial w_s} \right) p(\mathbf{x}, t) \, d\mathbf{x} \, dt$$
$$= \int \int y(\mathbf{x}, \mathbf{w}) \left(\frac{\partial^2 y(\mathbf{x}, \mathbf{w})}{\partial w_r \partial w_s} \right) p(\mathbf{x}) \, d\mathbf{x}$$

Simplifying the third term:

$$\int \int t \left(\frac{\partial^2 y(\mathbf{x}, \mathbf{w})}{\partial w_r \partial w_s} \right) p(\mathbf{x}, t) \, d\mathbf{x} \, dt$$

$$= \int \int t \left(\frac{\partial^2 y(\mathbf{x}, \mathbf{w})}{\partial w_r \partial w_s} \right) p(t|\mathbf{x})\, p(\mathbf{x})\, d\mathbf{x}\, dt$$

$$= \int \left(\frac{\partial^2 y(\mathbf{x}, \mathbf{w})}{\partial w_r \partial w_s} \right) \left(\int t\, p(t|\mathbf{x})\, dt \right) p(\mathbf{x})\, d\mathbf{x}$$

$$= \int \left(\frac{\partial^2 y(\mathbf{x}, \mathbf{w})}{\partial w_r \partial w_s} \right) \mathbb{E}[t|\mathbf{x}]\, p(\mathbf{x})\, d\mathbf{x}$$

From Section 1.5.5, we know that the optimal function that minimizes a sum-of-squares loss is the conditional average of the target data. This means that $y(\mathbf{x}, \mathbf{w}) = \mathbb{E}[t|\mathbf{x}]$.

Therefore, the second and third terms cancel out, and we get:

$$\frac{\partial^2 E}{\partial w_r \partial w_s} = \int \left(\frac{\partial y(\mathbf{x}, \mathbf{w})}{\partial w_r} \right) \left(\frac{\partial y(\mathbf{x}, \mathbf{w})}{\partial w_s} \right) p(\mathbf{x})\, d\mathbf{x}$$

which is the result in 5.194 that we wanted.