$$p(\mathbf{w}|X,\mathbf{y}) \propto \exp\big(-\frac{1}{2\sigma_n^2}(\mathbf{y}-X^\top\mathbf{w})^\top(\mathbf{y}-X^\top\mathbf{w})\big)\exp\big(-\frac{1}{2}\mathbf{w}^\top\Sigma_p^{-1}\mathbf{w}\big)$$
$$\propto \exp\big(-\frac{1}{2}(\mathbf{w}-\bar{\mathbf{w}})^\top\big(\frac{1}{\sigma_n^2}XX^\top+\Sigma_p^{-1}\big)(\mathbf{w}-\bar{\mathbf{w}})\big), \quad (2.7)$$

$$P\{w|X,y\} = P(y|X,w)\,p(w)\,/\,p(y|X)$$
$$\propto P(y|X,w)\cdot p(w)$$

since $y = X^T w + \varepsilon$ with $\varepsilon \sim N(0, \sigma_n^2)$

$$y \sim N(X^T w,\ \sigma_n^2 I)$$
$$w \sim N(0,\ \Sigma_p)$$

$\Rightarrow$

$$P(y|X,w) \propto \exp\{-\frac{1}{2\sigma_n^2}(y-X^Tw)^T(y-X^Tw)\}$$
$$p(w) \propto \exp\{-\frac{1}{2}w^T\Sigma_p^{-1}w\}$$

$$\Rightarrow\ P(w|X,y) \propto \exp\{-\frac{1}{2\sigma_n^2}(y-X^Tw)^T(y-X^Tw)\}\exp\{-\frac{1}{2}w^T\Sigma_p^{-1}w\}$$
$$= \exp\{-\frac{1}{2\sigma_n^2}[(y-X^Tw)^T(y-X^Tw)+\sigma_n^2 w^T\Sigma_p^{-1}w\}$$
$$= \exp\{-\frac{1}{2}(w-\hat{w})^T(\frac{1}{\sigma_n^2}XX^T+\Sigma_p^{-1})(w-\hat{w})\}$$

where $\hat{w} = \frac{1}{\sigma_n^2}(\frac{1}{\sigma_n^2}XX^T+\Sigma_p^{-1})Xy$

$$p(\mathbf{w}|X,\mathbf{y}) \sim \mathcal{N}(\bar{\mathbf{w}} = \frac{1}{\sigma_n^2}A^{-1}X\mathbf{y},\ A^{-1}), \quad (2.8)$$

Since $p(w|X,y) \propto \exp\{-\frac{1}{2}(w-\hat{w})^T\Sigma^{-1}(w-\hat{w})\}$

$$\Rightarrow\ w|X,y \sim N(\hat{w}, \Sigma)$$

s.t. $\begin{cases}\hat{w} = \frac{1}{\sigma_n^2}(\frac{1}{\sigma_n^2}XX^T+\Sigma_p^{-1})Xy \\ \Sigma = (\frac{1}{\sigma_n^2}XX^T+\Sigma_p^{-1})^{-1}\end{cases}$

$$
\begin{aligned}
p(f_*|\mathbf{x}_*, X, \mathbf{y}) &= \int p(f_*|\mathbf{x}_*, \mathbf{w})p(\mathbf{w}|X, \mathbf{y})\, d\mathbf{w} = \int \mathbf{x}_*^\top \mathbf{w}\; p(\mathbf{w}|X, \mathbf{y}) d\mathbf{w} \\
&= \mathcal{N}\Big(\frac{1}{\sigma_n^2}\mathbf{x}_*^\top A^{-1}X\mathbf{y},\; \mathbf{x}_*^\top A^{-1}\mathbf{x}_*\Big).
\end{aligned}
\tag{2.9}
$$

$$p(f_* | x_*, X, y) = \int p(f_* | x_*, w, X, y) \cdot p(w|X,y)\, dw$$

$$\propto \int x_*^T w \cdot \exp\{-\tfrac{1}{2}(w-\hat{w})^T \Sigma^{-1} (w-\hat{w})\}\, dw$$

correspond to $x_*^T E(w|X,y) = x_*^T \hat{w}$

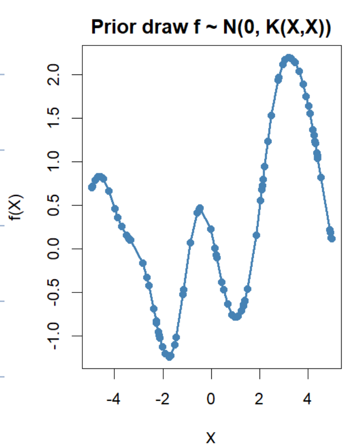
$\{f_* | x_*, X, y\}$ is a normal distribution of $p(w|X,y)$ scaled by x_*

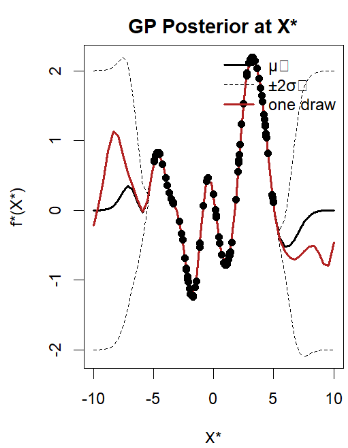
$$\Rightarrow N(x_*^T \hat{w} \quad x_*^T \Sigma x_*)$$

1. Replicate the generation of random functions from Figure 2.2. Use a regular (or random) grid of scalar inputs and the covariance function from eq. (2.16). Hints on how to generate random samples from multi-variate Gaussian distributions are given in section A.2. Invent some training data points, and make random draws from the resulting GP posterior using eq. (2.19).

$$\mathrm{cov}\big(f(\mathbf{x}_p), f(\mathbf{x}_q)\big) = k(\mathbf{x}_p, \mathbf{x}_q) = \exp\big(-\tfrac{1}{2}|\mathbf{x}_p - \mathbf{x}_q|^2\big). \quad \text{— (2.16)}$$

$$
\begin{aligned}
\mathbf{f}_*|X_*, X, \mathbf{f} \sim \mathcal{N}\big(&K(X_*, X)K(X, X)^{-1}\mathbf{f}, \\
&K(X_*, X_*) - K(X_*, X)K(X, X)^{-1}K(X, X_*)\big).
\end{aligned}
\tag{2.19}
$$





The left plot is a function drawn from Gaussian process given X as training data.

The right plot is the GP posterior using (2.19) given X^*.