Story

Observations

We have N i.i.d. scalar observation $\{x_n\} \subset \mathbb{R}$ that are sampled from some unknown $p(x_n)$. In our experiments, either x_n is sampled randomly at uniform,

$$x_n \sim \mathcal{U}_{[-1,1]},$$

or our dataset $\{x_n\}$ is a "linspace" over the range $[-1, 1]$,

$$x_n = -1 + \frac{2n}{N}.$$

Labels

We have

$$y = w^\top \phi(x) + \epsilon$$

where

$$\epsilon \sim \mathcal{N}(0, \sigma^2_{\text{out}}).$$

We call σ^2_{out} the output noise or observation noise. We call $\phi : \mathbb{R} \to \mathbb{R}^D$ our basis transformation and $w \in \mathbb{R}^D$ our weights. Our full dataset is $\mathcal{D} = \{(x_n, y_n)\}$

Weights

The weights have a prior Gaussian distribution

$$w \sim \mathcal{N}(\mu_w, \Sigma_w)$$

and a posterior Gaussian distribution

$$w \mid \mathcal{D} \sim \mathcal{N}(\hat{\mu}_w, \hat{\Sigma}_w)$$

where

$$\begin{aligned}
\hat{\Sigma}_w &= \left(2\Phi^{-1}\Phi + \Sigma_w^{-1}\right)^{-1}, \\
\hat{\mu}_w &= 2\hat{\Sigma}_w \Phi Y.
\end{aligned}$$

If we do not want a prior over weights, we set the prior covariance matrix to zero: $\Sigma_w = 0$.

Properties of Gaussians

Predictive marginalization cross entropy
"Cross-entropy" property

$$\begin{aligned}
\mathbb{E}_{x \sim \mathcal{N}(\mu_1, \Sigma_1)}[\mathcal{N}(x; \mu_2, \Sigma_2)] &= \int_x \mathcal{N}(x; \mu_1, \Sigma_1)\mathcal{N}(x; \mu_2, \Sigma_2)\mathrm{d}x \\
&= \mathcal{N}(\mu_1; \mu_2, \Sigma_1 + \Sigma_2) \\
&= \mathcal{N}(\mu_2; \mu_1, \Sigma_1 + \Sigma_2)
\end{aligned}$$