Problem 5:

$$\frac{1}{2}(\hat{\Phi} w - \hat{y})^T (\hat{\Phi} w - \hat{y}) = \frac{1}{2}\begin{pmatrix} \Phi w - y \\ \sqrt{\lambda} I_m w \end{pmatrix}^T \cdot \begin{pmatrix} \Phi w - y \\ \sqrt{\lambda} I_m w \end{pmatrix}$$

$$= \frac{1}{2}(\Phi w - y)^T (\Phi w - y) + \frac{1}{2}(\sqrt{\lambda} I_m w)^T \cdot (\sqrt{\lambda} I_m w)$$

$$= \frac{1}{2}\sum_{i=1}^{N} (\Phi_i w - y_i)^2 + \frac{\lambda}{2}\|w\|_2^2$$

ridge regression

Problem 6: a) with higher degree polynomial, the error always goes down.

b). use validation set, get w from training set and then test on validation set, choose the degree with lowest validation error.

Problem 7:

$$\log P(w, \beta | D) = \log P(w, \beta) + \log(y | \Phi, w, \beta) + \text{constant}$$

$$= \log N(w | m_0, \beta^{-1} S_0)\,\mathrm{Gamma}(\beta | a_0, b_0) + \log \prod_{i=1}^{N} N(y_i | w^T \phi(x_i), \beta)$$

$$= \log\left(\frac{\beta}{2\pi}\right)^{\frac{M}{2}} \exp\left(-\frac{\beta S_0^{-1}}{2}(w - m_0)^T (w - m_0)\right) + \sum_{i=1}^{N} \ln\left[\frac{\beta}{2\pi} \exp\left(-\frac{\beta}{2}(w^T \phi(x_i) - y_i)^2\right)\right]$$

$$+ (a_0 - 1)\log\beta - b_0\beta + a_0 \log b_0 - \log\Gamma(a_0)$$

$$= \frac{M}{2}\log\frac{\beta}{2\pi} - \frac{\beta S_0^{-1}}{2}(w - m_0)^T (w - m_0) + N \ln\frac{\beta}{2\pi} - \frac{\beta}{2}\sum_{i=1}^{N}(w^T \Phi(x_i) - y_i)^2 + (a_0 - 1)\log\beta$$

$$- b_0\beta + a_0 \log b_0 - \log\Gamma(a_0)$$

$$= \left(\frac{M}{2} + N\right)\log\frac{\beta}{2\pi} - \frac{\beta S_0^{-1}}{2}(w^T w - 2 m_0^T w + m_0^T m_0) - \frac{\beta}{2}(w^T \Phi^T \Phi w - 2 y^T \Phi w + y^T y) +$$

$$(a_0 - 1)\log\beta - b_0\beta + a_0 \log b_0 - \log\Gamma(a_0)$$

$$+ \text{constant}$$

$$-\frac{\beta S_0^{-1}}{2} m_0^T m_0$$

$$= \left(\frac{M}{2} + N + a_0 - 1\right)\log\beta + \beta S_0^{-1} m_0^T w - \frac{\beta S_0^{-1}}{2} w^T w - \frac{\beta \Phi^T \Phi}{2} w^T w + \beta y^T \Phi w - \frac{\beta}{2} y^T y - b_0\beta$$

$$+ a_0 \log b_0 - \log\Gamma(a_0)$$

$$+ \text{constant}$$

$$= \left(\frac{M}{2} + N + a_0 - 1\right)\log\beta + \left(-\frac{S_0}{2} m_0^T m_0 - \frac{y^T y}{2} - b_0\right)\beta + \left(\beta S_0^{-1} m_0^T + \beta y^T \Phi\right) w - \left(\frac{\quad}{2} + \frac{\quad}{2}\right) w^T w +$$

$$a_0 \log b_0 - \log\Gamma(a_0) + \text{constant}$$

$\log p(w, \beta \mid D)$

$$= -\frac{\beta S_N^{-1}}{2}(w - m_N)^T (w - m_N) + \frac{M}{2}\log\frac{\beta}{2\pi} + (a_N - 1)\log\beta - b_N\beta + a_N \log b_N - \log\Gamma(a_N)$$

$$= \left(\frac{M}{2} + a_N - 1\right)\log\beta + \left(-\frac{S_N^{-1}}{2} m_N^T m_N - b_N\right)\beta + \beta S_N^{-1} m_N^T w - \frac{\beta S_N^{-1}}{2} w^T w + a_N \log b_N -$$

$$\log\Gamma(a_N)$$

$$\frac{M}{2} + a_N - 1 = \frac{M}{2} + N + a_0 - 1$$

$$a_N = N + a_0$$

$$\begin{cases} -\frac{S_0^{-1}}{2} m_0^T m_0 - \frac{y^T y}{2} - b_0 = -\frac{S_N^{-1}}{2} m_N^T m_N - b_N \\ \beta S_0^{-1} m_0^T + \beta y^T \Phi = \beta S_N^{-1} m_N^T \\ \frac{\beta S_0^{-1}}{2} + \frac{\beta}{2}\Phi^T\Phi = \frac{\beta S_N^{-1}}{2} \end{cases}$$

$$S_0^{-1} + \Phi^T\Phi = S_N^{-1}$$

$$S_N = \left(S_0^{-1} + \Phi^T\Phi\right)^{-1}$$

$$\beta S_0^{-1} m_0^T + \beta y^T \Phi = \beta\left(S_0^{-1} + \Phi^T\Phi\right) m_N^T$$

$$m_N = \left(\left(S_0^{-1} + \Phi^T\Phi\right) \cdot \left(S_0^{-1} m_0^T + y^T \Phi\right)\right)^T$$

$$b_N = -\frac{S_N^{-1}}{2} m_N^T m_N + \frac{S_0^{-1}}{2} m_0^T m_0 + \frac{y^T y}{2} + b_0$$

$$a_N = N + a_0$$

Problem 8:

$$E_{\text{ridge}}(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{w}^T\phi(\boldsymbol{x}_i) - y_i)^2 + \frac{\lambda}{2}\boldsymbol{w}^T\boldsymbol{w}$$

$$E_{ridge}(w) = \frac{1}{2}\sum_{i=1}^{N}(w^T\Phi(x_i) - y_i)^2 + \frac{\lambda}{2}w^Tw$$

$$\nabla_w E_{ridge}(w) = \Phi^T(x)\Phi(x)\cdot w - \Phi^T(x)y + \lambda w$$

$$= (\Phi^T(x)\Phi(x) + \lambda\cdot I)w - \Phi^T(x)y$$

$$\nabla_w E_{ridge}(w) = 0$$

$$(\Phi^T(x)\Phi(x) + \lambda I)w = \Phi^T(x)y$$

$$w = (\Phi^T(x)\Phi(x) + \lambda I)^{-1}\Phi^T(x)y$$

if $N < M$, $\text{rank}(\Phi(x)) \le N$, $\text{rank}(\Phi^T(x)\Phi(x)) \le N$

$\Phi^T(x)\Phi(x) \in R^{M\times M}$ is not invertible.

introducing regularization can make $(\Phi^T(x)\Phi(x) + \lambda I)$ invertible.

Problem 9:

a).

$$\hat{y} = w^{*T}x$$

$$\hat{y} = w_{new}^T x_{new}$$

$$= w_{new}^T a x = w^{*T}x$$

$$w_{new} = \frac{w^*}{a}$$

b).

$$w^* = (x^Tx + \lambda I)^{-1}x^Ty$$

$$w_{new}^* = \frac{w^*}{a} = (x_{new}^T x_{new} + \lambda_{new} I)^{-1} x_{new}^T y$$

$$\frac{(X^TX+\lambda I)\ X^Ty}{a} = (a^2X+\lambda_{new} I)^{-1} a X^T y$$

$$(X^TX+\lambda I)^{-1} X^T y = (X + \frac{\lambda_{new}}{a^2} I)^{-1} X^T y$$

$$\lambda_{new} = a^2 \lambda$$