Linear Regression (w/ Matrix notation):

$$\text{model}\begin{cases} \underset{\sim}{y} \sim N(X\underset{\sim}{\beta}, \sigma^2 I) \\ \underset{\sim}{\beta} \sim N(\underset{\sim}{\mu}_\beta, \Sigma_\beta) \\ \sigma^2 \sim IG(q, r) \end{cases}$$

Note: $\underset{n\times 1}{\underset{\sim}{y}} = (y_1, \dots, y_n)'$

$\underset{n\times p}{X} = (\underset{\sim}{x}_0, \underset{\sim}{x}_1, \dots, \underset{\sim}{x}_{p-1})$

usually: $\underset{\sim}{x}_0 = \underset{\sim}{1}$

$$\underset{n\times n}{I} = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$

$\underset{p\times 1}{\underset{\sim}{\beta}} = (\beta_0, \beta_1, \dots, \beta_{p-1})'$

Posterior:

$$[\underset{\sim}{\beta}, \sigma^2 | \underset{\sim}{y}] \propto [\underset{\sim}{y} | \underset{\sim}{\beta}, \sigma^2][\underset{\sim}{\beta}][\sigma^2]$$

Full-conditional Distns:

$$[\underset{\sim}{\beta} | \cdot] \propto [\underset{\sim}{y} | \underset{\sim}{\beta}, \sigma^2][\underset{\sim}{\beta}]$$

$$\propto \exp\left\{-\frac{1}{2}(\underset{\sim}{y} - X\underset{\sim}{\beta})'(\sigma^2 I)^{-1}(\underset{\sim}{y} - X\underset{\sim}{\beta})\right\} \exp\left\{-\frac{1}{2}(\underset{\sim}{\beta} - \underset{\sim}{\mu}_\beta)'\Sigma_\beta^{-1}(\underset{\sim}{\beta} - \underset{\sim}{\mu}_\beta)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(-2\underset{\sim}{y}'(\sigma^2 I)^{-1}X\underset{\sim}{\beta} + \underset{\sim}{\beta}'X'(\sigma^2 I)^{-1}X\underset{\sim}{\beta} - 2\underset{\sim}{\mu}_\beta'\Sigma_\beta^{-1}\underset{\sim}{\beta} + \underset{\sim}{\beta}'\Sigma_\beta^{-1}\underset{\sim}{\beta}\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(-2\underbrace{\left(\underset{\sim}{y}'(\sigma^2 I)^{-1}X + \underset{\sim}{\mu}_\beta'\Sigma_\beta^{-1}\right)}_{\underset{\sim}{b}'}\underset{\sim}{\beta} + \underset{\sim}{\beta}'\underbrace{\left(X'(\sigma^2 I)^{-1}X + \Sigma_\beta^{-1}\right)}_{A}\underset{\sim}{\beta}\right)\right\}$$

$$= N(A^{-1}\underset{\sim}{b}, A^{-1})$$

$$[\sigma^2 | \cdot] \propto [\underset{\sim}{y} | \underset{\sim}{\beta}, \sigma^2][\sigma^2]$$

$$\propto |\sigma^2 I|^{-\frac{1}{2}} \exp\left\{-\frac{1}{\sigma^2}\frac{(\underset{\sim}{y} - X\underset{\sim}{\beta})'(\underset{\sim}{y} - X\underset{\sim}{\beta})}{2}\right\}(\sigma^2)^{-(q+1)} \exp\left\{\frac{-1}{\sigma^2 r}\right\}$$

$$\propto (\sigma^2)^{-(\underbrace{\frac{n}{2} + q}_{\tilde{q}} + 1)} \exp\left\{-\frac{1}{\sigma^2}\underbrace{\left(\frac{(\underset{\sim}{y} - X\underset{\sim}{\beta})'(\underset{\sim}{y} - X\underset{\sim}{\beta})}{2} + \frac{1}{r}\right)}_{\tilde{r}^{-1}}\right\} \propto IG(\tilde{q}, \tilde{r})$$

MCMC Algorithm:

1.) Initialize $\underline{\beta}^{(0)}$, $k=0$
2.) Let $k=k+1$
3.) Sample $(\sigma^2)^{(k)} \sim [\sigma^2 \mid \underline{\beta}^{(k-1)}, \underline{y}]$
4.) Sample $\underline{\beta}^{(k)} \sim [\underline{\beta} \mid (\sigma^2)^{(k)}, \underline{y}]$
5.) Goto 2 for $k=1,\ldots,K$

Prediction:

Bayesian prediction involves the posterior predictive distn (PPD):

$$[\tilde{y} \mid \underline{y}] = \int [\tilde{y}, \underline{\theta} \mid \underline{y}]\, d\underline{\theta}$$

$$= \int \underbrace{[\tilde{y} \mid \underline{\theta}, \underline{y}]}_{\substack{\text{predictive} \\ \text{full-cond.} \\ \text{distn}}} \underbrace{[\underline{\theta} \mid \underline{y}]}_{\substack{\text{posterior} \\ \text{distn}}}\, d\underline{\theta}$$

unobserved data

Steps: 1.) derive pred. full-cond. distn
2.) Use composition sampling:
a.) $\underline{\theta}^{(k)} \sim [\underline{\theta} \mid \underline{y}]$, for $k=1,\ldots,K$
b.) $\tilde{y} \sim [\tilde{y} \mid \underline{\theta}^{(k)}, \underline{y}]$

3.) use MC integration:

$$E(\tilde{y} \mid \underline{y}) = \frac{\sum_{k=1}^{K} \tilde{y}^{(k)}}{K}$$ point prediction

Regression PPD:

Note: $[\tilde{y} \mid \beta, \sigma^2, y] = [\tilde{y} \mid \beta, \sigma^2]$ b/c $\tilde{y}$ and y are conditionally independent

For $k = 1, \ldots, K$, sample:

$$\tilde{y}^{(k)} \sim N(\tilde{X}\beta^{(k)}, \sigma^{2(k)} I),$$

$\uparrow$ design matrix for predictions

$$E(\tilde{y} \mid y) = \frac{\sum_{k=1}^{K} \tilde{y}^{(k)}}{K}$$

posterior point predictions

Note: This is different than fitted values:

$$E(X\beta \mid y) = \frac{\sum_{k=1}^{K} X\beta^{(k)}}{K}$$