1. Assume $y \sim N(X\beta, s^2\mathrm{I})$ where $X$ is $n \times p$.

   (a) The prior distribution whose MAP corresponds to the ridge estimator is

$$\beta \sim N(0, \lambda^{-1}\mathrm{I}_p)$$

In this prior, $\lambda$ controls the amount of shrinkage. The variance of the prior is solely dependant on $\lambda$ so as it increase, the variance decreases and becomes tighter around the mean(in this case, 0).

Putting it together we get:

$$f(y|\ldots) \propto exp\{-\frac{1}{2s^2}(y - X\beta)^T(y - X\beta)\}$$
$$p(\beta|\lambda) \propto exp\{-\frac{\lambda}{2}\beta^\top\beta\}$$
$$p(\beta|\ldots) \propto exp\{-\frac{1}{2s^2}(y - X\beta)^\top(y - X\beta) - \frac{\lambda}{2}\beta^\top\beta\}$$

Then to get the MAP, we take the maximum:

$$\hat{\beta}_R = \text{argmax}_\beta \left[exp\{-\frac{1}{2s^2}(y - X\beta)^\top(y - X\beta) - \frac{\lambda}{2}\beta^\top\beta\}\right]$$
$$\hat{\beta}_R = \text{argmin}_\beta \left[\frac{1}{2s^2}(y - X\beta)^\top(y - X\beta) + \frac{\lambda}{2}\beta^\top\beta\right]$$
$$\hat{\beta}_R = \text{argmin}_\beta \left[(y - X\beta)^\top(y - X\beta)/2 + \lambda'\beta^\top\beta\right]$$

   (b) The prior distributions whose MAP corresponds to the LASSO estimator are

$$p(\beta|\sigma^2) \propto \prod_{j=1}^{p} \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta_j|/\sqrt{\sigma^2}}$$
$$p(\sigma^2) \propto \frac{1}{\sigma^2}$$

In this prior, $\lambda$ also controls the amount of shrinkage, but it is not as straight forward as the cases before. Here we have a Laplace prior with variance $2\frac{\sigma^2}{\lambda^2}$. So as $\lambda$ increases the variance will go to 0. So $\lambda$ controls again the shrinkage, but in this case $\sigma^2$ will also affect the how much $\lambda$ influences shrinkage.

Putting it together we get:

$$f(y|\ldots) \propto exp\{-\frac{1}{2s^2}(y - X\beta)^T(y - X\beta)\}$$
$$p(\beta|\sigma^2, \lambda) \propto exp\{-\lambda \sum_{j=1}^{p} |\beta_j|/\sqrt{\sigma^2}\}$$
$$p(\beta|\ldots) \propto exp\{-\frac{1}{2s^2}(y - X\beta)^\top(y - X\beta) - \lambda \sum_{j=1}^{p} |\beta_j|/\sqrt{\sigma^2}\}$$

Then to get the MAP, we take the maximum:

$$\hat{\beta}_R = \text{argmax}_\beta \left[exp\{-\frac{1}{2s^2}(y - X\beta)^\top(y - X\beta) - \lambda \sum_{j=1}^{p} |\beta_j|/\sqrt{\sigma^2}\}\right]$$
$$\hat{\beta}_R = \text{argmin}_\beta \left[\frac{1}{2s^2}(y - X\beta)^\top(y - X\beta) + \frac{\lambda}{\sqrt{\sigma^2}} \sum_{j=1}^{p} |\beta_j|\right]$$
$$\hat{\beta}_L = \text{argmin}_\beta \left[(y - X\beta)^\top(y - X\beta)/2 + \lambda' \sum_{j=1}^{p} |\beta_j|\right]$$

(c) The elastic net estimator is

$$p(\beta|\sigma^2) \propto exp\{-\frac{1}{2\sigma^2}(\lambda_1||\beta||_1 + \lambda_2||\beta||_2^2)\}$$
$$p(\sigma^2) \propto \frac{1}{\sigma^2}$$

The $\lambda$ parameters control the blend between a Normal and Laplace distribution. When $\lambda_2 > \lambda_1$, the prior takes more of a form of a normal distribution which corresponds to the prior for the ridge estimator and when $\lambda_1 > \lambda_2$, the prior takes more of a form of a Laplace distribution which corresponds to the prior for the LASSO estimator.

Putting it together we get:

$$f(y|\ldots) \propto exp\{-\frac{1}{2s^2}(y - X\beta)^T(y - X\beta)\}$$
$$p(\beta|\ldots) \propto exp\{-\frac{1}{2\sigma^2}(\lambda_1||\beta||_1 + \lambda_2||\beta||_2^2)\}$$
$$p(\beta|\ldots) \propto exp\{-\frac{1}{2s^2}(y - X\beta)^\top(y - X\beta) - \frac{1}{2\sigma^2}(\lambda_1||\beta||_1 + \lambda_2||\beta||_2^2)\}$$

Then to get the MAP, we take the maximum:

$$\hat{\beta}_R = \text{argmax}_\beta \left[exp\{-\frac{1}{2s^2}(y - X\beta)^\top(y - X\beta) - \frac{1}{2\sigma^2}(\lambda_1||\beta||_1 + \lambda_2||\beta||_2^2)\}\right]$$
$$\hat{\beta}_R = \text{argmin}_\beta \left[\frac{1}{2s^2}(y - X\beta)^\top(y - X\beta) + \frac{1}{2\sigma^2}(\lambda_1||\beta||_1 + \lambda_2||\beta||_2^2)\right]$$
$$\hat{\beta}_E = \text{argmin}_\beta \left[(y - X\beta)^\top(y - X\beta)/2 + \lambda_1\beta'\beta + \lambda_2\sum_{j=1}^{p}|\beta_j|\right]$$