Maximum Likelihood Estimate

- Suppose that true data generated from probabilistic model with Gaussian noise:

$$\boldsymbol{y} = A\boldsymbol{\beta} + \boldsymbol{w}, \qquad w_i \sim N(0, \sigma^2)$$

- Maximum likelihood estimator:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\beta} p(\boldsymbol{y}|A, \boldsymbol{\beta}) = \arg\min_{\beta} [-\ln p(y|A, \boldsymbol{\beta})]$$

- Gaussian density for noise in y: $\ln p(\boldsymbol{y}|A, \boldsymbol{\beta}) = -\frac{1}{2\sigma^2} \|\boldsymbol{y} - A\boldsymbol{\beta}\|^2$

- Hence

$$\hat{\boldsymbol{\beta}} = \arg\max_{\beta} p(\boldsymbol{y}|A, \boldsymbol{\beta}) = \arg\min_{\beta} [\|\boldsymbol{y} - A\boldsymbol{\beta}\|^2] = \text{Least Squares Solution}$$

Bayes Estimation (MAP Estimate)

- Maximum a posterior (MAP) estimator of $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\beta}} = \arg max_{\beta}\, p(\boldsymbol{\beta}|\boldsymbol{y}, A)$$

  - $\hat{\beta}$ = Most likely parameter value given evidence y, A

- Bayes Rule: $p(\boldsymbol{\beta}\,|\boldsymbol{y}, A) = p(\boldsymbol{y}|A, \boldsymbol{\beta}\,)p(\boldsymbol{\beta})/p(\boldsymbol{y}|A)$

- Hence: $\widehat{\boldsymbol{\beta}} = \arg\max_{\beta} p(y|A, \boldsymbol{\beta}\,)p(\boldsymbol{\beta})$ (because $\boldsymbol{y}$ and A are fixed)
  - Likelihood: $p(y|A, \boldsymbol{\beta}\,)$ How well $\boldsymbol{\beta}$ matches data
  - Prior: $p(\boldsymbol{\beta}\,)$: How well $\boldsymbol{\beta}$ agrees with prior knowledge about its distribution (constraints)

- More in probability class…

Bayes Estimation with Logarithms

- Often easier to use logarithms:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\beta} p(\boldsymbol{y}|A,\boldsymbol{\beta})p(\boldsymbol{\beta}) = \arg\min_{\beta}[-\ln p(\boldsymbol{y}|A,\boldsymbol{\beta})p(\boldsymbol{\beta})]$$
$$= \arg\min_{\beta}[-\ln p(\boldsymbol{y}|A,\boldsymbol{\beta}) - \ln p(\boldsymbol{\beta})]$$

- Gaussian density for noise in y: $\ln p(\boldsymbol{y}|A,\boldsymbol{\beta}) = -\frac{1}{2\sigma^2}\|\boldsymbol{y} - A\boldsymbol{\beta}\|^2$

- Hence

$$\hat{\beta} = \arg\min_{\beta}\left[\frac{1}{2\sigma^2}\|\boldsymbol{y} - A\boldsymbol{\beta}\|^2 - \ln p(\boldsymbol{\beta})\right] = \arg\min_{\beta}[\|\boldsymbol{y} - A\boldsymbol{\beta}\|^2 + \phi(\boldsymbol{\beta})]$$

- Conclusion: MAP estimate = regularized LS with $\phi(\boldsymbol{\beta}) = -2\sigma^2 \ln p(\boldsymbol{\beta})$
  - Penalize $\boldsymbol{\beta}$ proportional to $-\ln p(\boldsymbol{\beta})$: Less likely $\boldsymbol{\beta}$ penalized more

Ridge and Lasso as Bayesian Estimators

- Bayesian Estimator:

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\beta}\left[\frac{1}{2\sigma^2}\|\boldsymbol{y} - A\,\boldsymbol{\beta}\|^2 - \ln p(\boldsymbol{\beta})\right]$$

- Assuming β_j are i.i.d. Gaussian with zero mean:

$$p(\beta_j) = \tfrac{1}{2\pi\sigma} exp(-\beta_j{}^2/2\gamma^2),\ -\log p(\beta_j) = \beta_j{}^2/2\gamma^2 + constants$$

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\beta}\left[\|\boldsymbol{y} - A\boldsymbol{\beta}\|^2 + \tfrac{\sigma^2}{\gamma^2}\|\,\boldsymbol{\beta}\|^2\right] \quad = \text{Ridge Regression!}$$

- Assuming β_j are i.i.d. Laplacian with zero mean:

$$p(\beta_j) = \tfrac{1}{2\sigma} exp(-|\beta_j|/\gamma\,),\ -\log p(\beta_j) = |\beta_j|/\gamma\ + constant$$

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\beta}\left[\|\boldsymbol{y} - A\boldsymbol{\beta}\,\|^2 + \tfrac{2\sigma^2}{\gamma}\|\boldsymbol{\beta}\|_1\right] \quad = \text{Lasso Regression!}$$