BIOS 526 Course Summary

1. Ordinary least squares (M1):

$$\mathrm{Cov}\,\mathbf{Y} = \sigma^2\mathbf{I}$$
$$\hat{\boldsymbol{\beta}}^{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

2. Generalized least squares (M4, slide 16) is a formulation for any given covariance matrix:

$$\mathrm{Cov}\,\mathbf{Y} = \boldsymbol{\Sigma}$$
$$\hat{\boldsymbol{\beta}}^{GLS} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}$$

   - Can apply to clustered data; in practice, use LMMs or GEEs.
   - Can be used for heteroscedastic data; in practice, use GEE.
   - Need to know $\boldsymbol{\Sigma}$; in practice, estimated.

3. Generalized linear models (M3 part I):

$$y_i \overset{ind}{\sim} P(\boldsymbol{x}_i'\boldsymbol{\beta})$$
$$g(E(y_i)) = \boldsymbol{x}_i'\boldsymbol{\beta}$$

   - logistic regression for 0, 1 response. $\mathrm{Var}\,(y_i) = E(y_i)\,\{1 - E(y_i)\}$.
   - Poisson regression for count data. $\mathrm{Var}\,(y_i) = E(y_i)$. Watch out for overdispersion.

4. Generalized linear mixed models (M2, M3 part II):

   Random intercept model:

$$y_{ij} \sim P(\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \theta_i)$$
$$g\,\{E(y_{ij}|\theta_i)\} = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + \theta_i$$
$$\theta_i \overset{iid}{\sim} N(0, \tau^2)$$

- Handle repeated measurements / longitudinal / clustered data.
- For Gaussian, interpretation of $\boldsymbol{\beta}$ not impacted by conditional versus marginal (the estimates of $\boldsymbol{\beta}$ from LMM and GEE are different but usually similar, in some cases GEE with exchangeable correlation structure and random intercept LMM have equivalent $\hat{\boldsymbol{\beta}}$).
- For logistic, interpretation of $\boldsymbol{\beta}$ in GLMM (conditional model) is different from the interpretation in a GEE (the estimates of $\boldsymbol{\beta}$ from GLMM and GEE are different).
- Use if interested in subject-specific predictions (shrinkage towards population effects).
- Can use if no overdispersion in logistic or Poisson, no heteroscedasticity in Gaussian.

5. Generalized estimating equations (M4):

$$y_{ij} \sim P(\boldsymbol{x}'_{ij}\boldsymbol{\beta})$$

$$\text{Cov}\,(\boldsymbol{y}_i) = \mathbf{D}_i^{1/2}\mathbf{R}(\alpha)\mathbf{D}_i^{1/2}$$

where $\mathbf{R}(\alpha)$ is the working correlation and $\mathbf{D}_i$ is a diagonal matrix with diagonal elements equal to the variance determined by the likelihood.

- Handle repeated measurements / longitudinal / clustered data.
- Use robust standard errors.
- Use if heteroscedasticity and/or overdispersion (valid inference, unlike GLMM).
- Marginal inference (no random effects).

6. Generalized additive models (M5):

$$g(E(y_i)) = \beta_0 + s_1(x_{i1}) + \cdots + s_j(x_{ip})$$

- Handle non-linear effects.
- Can incorporate random effects for longitudinal / repeated measures / clustered data.
- Can generalize interactions from linear models to bivariate splines, e.g., $s(x_{i1}, x_{i2})$, i.e., 2D surfaces.
- Estimate $s(x_{ip})$ using either cross-validation or mixed model formulation of spline coefficients.

7. Bias-Variance Tradeoff (M5, slides 33-43, M6 II slides 5-6)

- $MSE(\hat{f}(x)) = Var(\hat{f}(x)) + Bias(\hat{f}(x))^2$
- Fewer parameter: more bias, less variance

- More parameters: less bias, more variance
- Use cross-validation or generalized cross-validation to approximately minimize the MSE

8. Principal component analysis (M6 I): uses the singular value decomposition on standardized data:
$$\mathbf{X}_{scaled} = \mathbf{UDV}'$$
   - Lower dimensional representation using first l left eigenvectors.
   - Can use in principal component regression when have issues with multicollinearity.

9. Ridge Regression (L2-norm regularization) (M5 part II, M6 part II):
$$\hat{\boldsymbol{\beta}}^{Ridge} = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}$$
   - Regularization method with a nice closed form.
   - Can use when lots of covariates, $p > n$.
   - Use for shrinking spline coefficients in GAMs (used in MGCV).

10. Lasso (L1-norm regularization) (M6 part II):
$$\hat{\boldsymbol{\beta}}^{Lasso} = \underset{\boldsymbol{\beta}}{\text{argmin}} \quad -\sum_{i=1}^{n} \ell(y_i; \boldsymbol{x}_i'\boldsymbol{\beta}) + \lambda||\boldsymbol{\beta}||_1$$
    - Regularization that results in variable selection by setting many coefficients equal to 0.

11. Elastic net (L1-norm and L2-norm regularization) (M6 part II):
$$\hat{\boldsymbol{\beta}}^{ElNet} = \underset{\boldsymbol{\beta}}{\text{argmin}} \quad -\sum_{i=1}^{n} \ell(y_i; \boldsymbol{x}_i'\boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} \left(\alpha|\beta_j| + \frac{(1-\alpha)}{2}\beta_j^2 \right).$$
    - A good choice when predictors are correlated.
    - Use for variable selection.

12. Bayesian statistics (M7):

(a) Choose a prior $P(\boldsymbol{\theta})$.

(b) Choose a likelihood $P(\mathbf{Y}|\boldsymbol{\theta})$.

(c) Collect data and calculate posterior $P(\boldsymbol{\theta}|\mathbf{Y})$.

- Connection between regularized regression and priors on the parameters.
- In Bayesian regression with a Gaussian likelihood and a Gaussian prior, equivalent to ridge regression
- In Bayesian regression with a Laplace (i.e., double exponential) prior, the maximum a posteriori estimate corresponds to the Lasso solution.
- Parameters are random variables.
- In practice, obtain samples from the posterior distribution.
- Calculate credible intervals from quantiles of a posterior sample.