DS6410: Homework 2

Due: Mar 2, 2026

Q1. Fourier Basis Expansion for Image Modeling. In this problem, you will study *Fourier basis expansion* as a method for representing and learning nonlinear structure in imaging data, combining theoretical understanding with practical implementation.

Let $I(u, v)$ denote a grayscale image of size $H \times W$, where (u, v) indexes pixel location. Treat the image as a function

$$f(u, v) = I(u, v).$$

Consider the following real-valued 2D Fourier basis functions:

$$\phi_{k\ell}^{(c)}(u, v) = \cos(2\pi k u)\cos(2\pi \ell v), \qquad \phi_{k\ell}^{(s)}(u, v) = \sin(2\pi k u)\sin(2\pi \ell v),$$

for integers $k, \ell \geq 0$.

1. Explain how the collection of functions

$$\left\{\phi_{k\ell}^{(c)}(u, v),\ \phi_{k\ell}^{(s)}(u, v)\right\}_{k,\ell}$$

   defines a *basis expansion* for functions on $[0, 1]^2$.

2. Consider the truncated Fourier expansion

$$f_K(u, v) = \sum_{k=0}^{K}\sum_{\ell=0}^{K}\Big[a_{k\ell}\cos(2\pi k u)\cos(2\pi \ell v) + b_{k\ell}\sin(2\pi k u)\sin(2\pi \ell v)\Big].$$

   (a) How does the truncation level K control model complexity?

   (b) Discuss the bias–variance tradeoff as K increases.

   (c) Why does overfitting become a concern for large K?

3. Coding Part.

   (a) Constructing Fourier features:

      i. Load a grayscale image and normalize pixel coordinates to $[0, 1]^2$. (You may use any standard grayscale image.)

ii. Sample n pixel locations $\{(u_i, v_i)\}_{i=1}^n$ and form responses $y_i = I(u_i, v_i)$.
iii. For a given K, construct the design matrix Φ_K using a truncated 2D Fourier basis.

(b) Fitting and regularization:

i. Fit a linear model using Φ_K via:
   - ordinary least squares (when $p_K \leq n$),
   - ridge regression (choose λ by cross-validation).

ii. For several values of K, report training and test mean squared error.

(c) Reconstruction and visualization

i. Use the fitted model to predict intensities at all pixel locations.
ii. Reconstruct and display the estimated image $\widehat{I}_K$ for at least three values of K.

Q2. Ridge Regression as Bayesian MAP Estimation. Consider the linear model

$$y = X\beta + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2 I_n),$$

where $y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n \times p}$, and $\beta \in \mathbb{R}^p$.

1. **Gaussian prior.** Assume a prior distribution on the regression coefficients

$$\beta \sim \mathcal{N}(0, \tau^2 I_p).$$

Derive the **posterior distribution** $p(\beta \mid y)$, and give its mean and covariance in closed form.

2. **MAP estimator.** Derive the maximum a posteriori (MAP) estimator

$$\hat{\beta}_{\mathrm{MAP}} := \arg\max_{\beta} p(\beta \mid y),$$

and show that it is the unique minimizer of an optimization problem of the form

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_2^2.$$

3. **Ridge connection.** Conclude that $\hat{\beta}_{\mathrm{MAP}}$ coincides with the ridge regression estimator, and identify the relationship between the ridge penalty parameter λ and the variance parameters σ^2 and τ^2.

Hint: Write the log posterior $\log p(\beta \mid y)$ up to an additive constant and compare it to a penalized least squares objective.

Q3. Coding: Group Lasso on `bardet`. In this problem, you will fit and interpret a *group lasso* model on an established dataset. Bardet is gene expression data (20 genes for 120 samples) from the microarray experiments of mammalian eye tissue samples of Scheetz et al. (2006).

Goal. Compare **group lasso** against **lasso** and **ridge** in terms of predictive performance and interpretability.

Dataset. Use the dataset `bardet`:

1. **Load and construct (X, y) and groups.**
   (a) Load the data with `data(bardet)` and inspect the resulting object(s). Identify the response vector y and the predictor matrix X (or construct X from a data.frame).
   (b) Define a *group membership vector* $g \in \{1, \dots, G\}^p$ for the p columns of X. Each 5 consecutive columns to a grouped gene.

2. **Train/test split and standardization.** Create a train/test split (e.g., 70/30) using a fixed random seed. Standardize the columns of X using *training-set* means and standard deviations, and apply the same transformation to the test set.

3. **Fit group lasso and tune λ.** Fit a group lasso model on the training data by solving

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2n_{\text{tr}}} \|y_{\text{tr}} - X_{\text{tr}}\beta\|_2^2 \; + \; \lambda \sum_{k=1}^{G} w_k \|\beta_{(k)}\|_2,$$

   where $\beta_{(k)}$ denotes the subvector of coefficients in group k and w_k are optional group weights (e.g., $w_k = \sqrt{|k|}$). Choose λ using K-fold cross-validation on the training set.
   Report: the selected λ and a plot of CV error versus $\log(\lambda)$.

4. **Fit comparison models (lasso and ridge).** Fit a lasso model and a ridge regression model on the same training data, selecting λ via cross-validation. **Report:** the selected λ for each method, and a plot of CV error versus $\log(\lambda)$ for each method (overlay is fine).

5. **Evaluate predictive performance.** Evaluate each fitted model on the test set using an appropriate metric. **Report:** the test-set metric for group lasso, lasso, and ridge in a small table.

6. **Interpretation: groups vs individual predictors.**
   (a) For group lasso, report which *groups* are selected (i.e., groups with $\|\hat{\beta}_{(k)}\|_2 > 0$).
   (b) For lasso, report which *individual* predictors are selected (nonzero coefficients).

(c) Compare sparsity patterns and interpretability: does group lasso yield a more interpretable model for `bardet`? Explain briefly with any supporting output/plots.

7. **Stability check.** Repeat Steps 2–6 for three different random seeds. Comment on the stability of selected groups/features across splits.

Submission: Submit your code (well-documented) and a short write-up containing the requested plots, tables, and discussion.