# SDS 383D: Exercises 3 – Linear smoothing and Gaussian processes

February 26, 2017

*Professor Scott*

**Spencer Woody**

# Problem 1

## Basic Concepts

(A)

# Problem 2

## Curve fitting by linear smoothing

In this problem, consider a general nonlinear regression with one predictor and one response, $y_i = f(x_i) + \epsilon_i$, where $\epsilon_i$ are mean-zero random variables.

(A) For now, consider a linear regression on a response $y_i$ with one predictor $x_i$, and both $y_i$ and $x_i$ have had their means subtracted, so the $y_i = \beta x_i + \epsilon_i$. Define $S_x := \sum_{i=1}^n x_i^2$. The least squares estimate for the coefficient, from Exercises 1, is

$$\begin{aligned}\hat{\beta} &= (X^T X)^{-1} X^T y \\ &= (x^T x)^{-1} x^T y \\ &= \frac{\sum_{i=1}^n x_i \cdot y_i}{\sum_{i=1}^n x_i^2} \\ &= \frac{\sum_{i=1}^n x_i \cdot y_i}{S_x} \\ &= \sum_{i=1}^n \frac{x_i}{S_x} \cdot y_i.\end{aligned}$$

So now our prediction $y^\star | x^\star$ is,

$$\begin{aligned}\hat{y}^\star &= \hat{f}(x^\star) \\ &= \hat{\beta} x^\star \\ &= \left( \sum_{i=1}^n \frac{x_i}{S_x} \cdot y_i \right) \cdot x^\star \\ &= \sum_{i=1}^n \left( \frac{x_i}{S_x} \cdot x^\star \right) \cdot y_i,\end{aligned}$$

which we recognize as being in the form of the general *linear smoother*

$$\hat{f}\,(x^\star) = \sum_{i=1}^n w(x_i, x^\star) \cdot y_i$$

for some weight function $w(x_i, x^\star)$. In particular, the weight function for linear regression gives weight to each $y_i$ proportional to the value of $x_i$. Contrast this with the $k$-nearest neighbors smoothing weight function,

$$w_K(x_i, x^\star) = \begin{cases} 1/K & \text{if } x_i \text{ is one of the K closest sample points to } x^\star \\ 0 & \text{otherwise} \end{cases},$$

which gives *equal* weight to $y_i$s but *only* to the $k$-nearest neighbors of $x^\star$.

(B) Now we have the very general weight function

$$w(x_i, x^\star) = \frac{1}{h} \cdot K\left( \frac{x_i - x^\star}{h}, \right)$$

where $K(\bullet)$ is some kernel function. The script `myfuns03.R` in the appendix shows an R function for linear smoothing, as well functions for the uniform and Gaussian kernels. Figure 1 shows an example of smoothing with a bandwidth of 0.75 for a cubic function $f(x)$ with iid residuals from the $\mathcal{N}(0, 15^2)$ distribution.

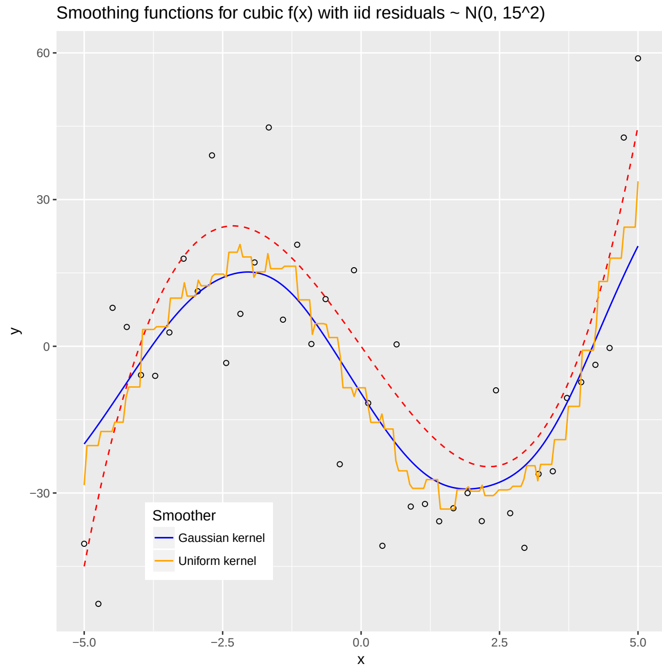


Figure 1: Uniform and Gaussian kernel smoothing for $y = f(x) + e$, $f(x) = x(x-4)(x+4)$, $h = 0.75$

# Problem 3

## Cross validation

(A) See attached R code for a script to return prediction error estimates for smoothing given a specified choice of bandwidth, $h$.

(B) For this exercise, I produced 500 data points on the $x$-space $[0, 1]$ from a sine function $f(x)$ with a given period and set the amplitude, and added Gaussian noise with a given standard deviation. Then I used 5-fold cross validation to select the optimal bandwidth for that given period and standard deviation of noise term. Figure 2 shows the optimal bandwidths for period ranging from 0.1 to 1, and standard deviation ranging from 0.001 to 0.5, and Figure 3 shows four example The highest bandwidths are chosen for functions with high "wigglyness" and high noise, and the smallest bandwidths are chosen for functions with low "wigglyness" and low noise. This makes sense. As the frequency increases (i.e., period decreases) then we need a tighter bandwidth because the value of the function is fluctuating at a greater rate. As noise increases, we need a greater bandwidth to smooth out the noise. Furthermore, we can see that in all cases we recover the underlying function pretty well.

(C) I'll get around to this eventually. . . .

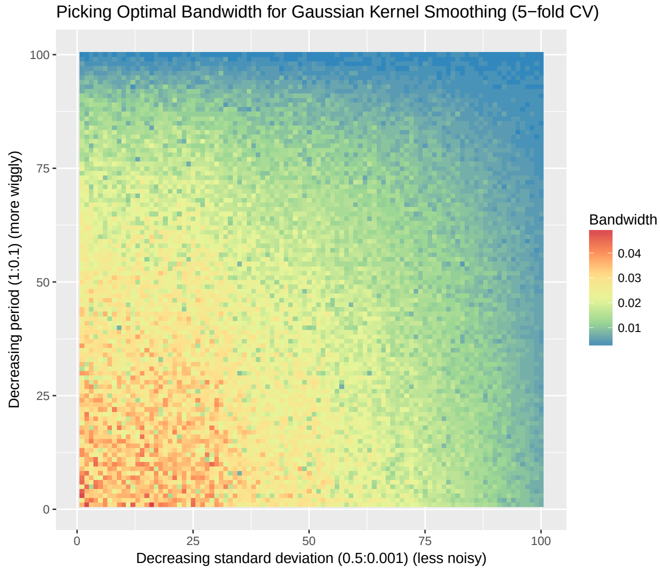


Figure 2: Optimal bandwidths for varying periods and standard deviations

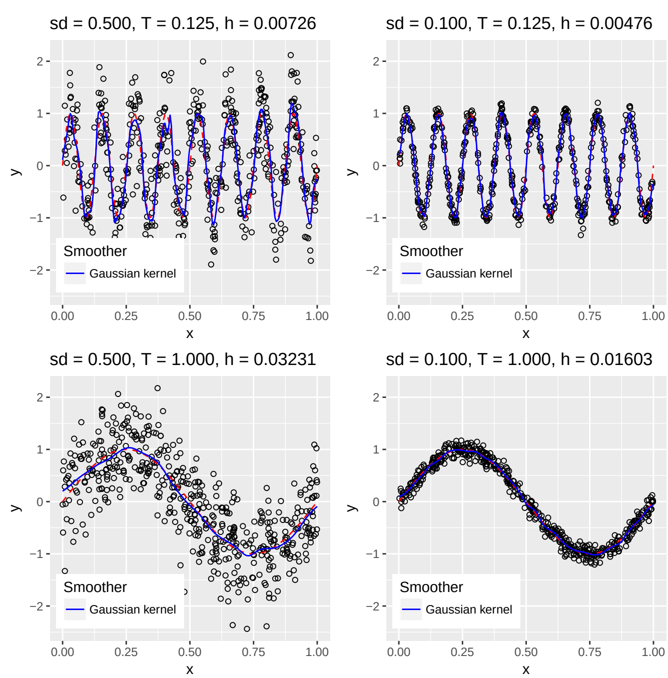


Figure 3: $2 \times 2$ example with fitted curves

# Problem 4

## Local polynomial regression

(A) Define

$$
\begin{aligned}
g_x(u; a) &= a_0 + \sum_{k=1}^{D} a_k (u - x)^k \\
&= \begin{cases} \sum_{j=0}^{D+1} a_j (u-x)^j & \text{if } u \neq x \\ a_0 & \text{if } u = x \end{cases}.
\end{aligned}
$$

The coefficients of $a$ for the local polynomial regression with dimension $D$ will come from the weighted least squares problem

$$
\hat{a} = \arg \min_{a \in \mathcal{R}^{D+1}} \sum_{i=1}^{n} w_i \left[ y_i - g_x(x_i, a) \right]^2 \tag{1}
$$

Furthermore, define the matrix $R_x$ whose $(i, j)$ element is $(x_i - x)^2$. Then the estimate $\hat{f}(x)$ will be $R_x \hat{a}$. The solution of $\hat{a}$ may be found with

$$
\begin{aligned}
\hat{a} &= \arg \min_{a \in \mathcal{R}^{D+1}} (y - R_x a)^T W (y - R_x a) \\
&= (R_x^T W R_x)^{-1} R_x^T W y
\end{aligned}
$$

where $W = \text{diag}(w_1, \ldots, w_n)$ is a diagonal matrix of weights from some kernel,

$$
w_i = \frac{1}{h} K \left( \frac{x_i - x}{h} \right)
$$

and this solution is found following the same argument to find the WLS estimate of linear regression from Exercises 1.

(B) Define the matrix $B_x = (R_x^T W R_x)^{-1} R_x^T W$. The estimate of $f$ at a point $x^\star$ is $\hat{f}(x) = \hat{a}_0$, the first element of the vector $\hat{a}$ whose form is derived above. Since our estimate is in the form of a linear smoother, this can also be written as $\hat{f}(x) = \frac{b_x^T y}{\sum_{i=1}^n b_{x,i}}$ if we take $b^T$ to be the first row of $B$. In the special case of the local linear smoother ($D = 1$), the matrix $R_x$ has dimension $n \times 2$ and can be written as

$$
R_x = \begin{bmatrix} 1 & x_1 - x \\ \vdots & \vdots \\ 1 & x_n - x \end{bmatrix}.
$$

$$
R_x^T = \begin{bmatrix} 1 & \ldots & 1 \\ x_1 - x & \ldots & x_n - x \end{bmatrix}
$$

$$
W = \begin{bmatrix} w_1 & & \mathcal{O} \\ & \ddots & \\ \mathcal{O} & & w_n \end{bmatrix}
$$

$$
R_x^T W R_x = \begin{bmatrix} \sum_{i=1}^n w_i & \sum_{i=1}^n w_i (x_i - x) \\ \sum_{i=1}^n w_i (x_i - x) & \sum_{i=1}^n w_i (x_i - x)^2 \end{bmatrix}^{-1}
$$

$$R_x^T W = \begin{bmatrix} w_1 & \dots & w_n \\ w_1(x_1 - x) & \dots & w_n(x_n - x) \end{bmatrix}$$

Furthermore, define the term

$$s_j(x) = \sum_{i=1}^{n} w_i (x_i - x)^j.$$

Now we can find the exact form of $B$ up to a proportionality constant,

$$\begin{aligned}
B &= (R_x^T W R_x)^{-1} R_x^T W \\
&= \left( \begin{bmatrix} 1 & \dots & 1 \\ x_1 - x & \dots & x_n - x \end{bmatrix} \begin{bmatrix} w_1 & & \mathcal{O} \\ & \ddots & \\ \mathcal{O} & & w_n \end{bmatrix} \begin{bmatrix} 1 & x_1 - x \\ \vdots & \vdots \\ 1 & x_n - x \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & \dots & 1 \\ x_1 - x & \dots & x_n - x \end{bmatrix} \begin{bmatrix} w_1 & & \mathcal{O} \\ & \ddots & \\ \mathcal{O} & & w_n \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^n w_i & \sum_{i=1}^n w_i(x_i - x) \\ \sum_{i=1}^n w_i(x_i - x) & \sum_{i=1}^n w_i(x_i - x)^2 \end{bmatrix}^{-1} \begin{bmatrix} w_1 & \dots & w_n \\ w_1(x_1 - x) & \dots & w_n(x_n - x) \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^n w_i & s_1(x) \\ s_1(x) & s_2(x) \end{bmatrix}^{-1} \begin{bmatrix} w_1 & \dots & w_n \\ w_1(x_1 - x) & \dots & w_n(x_n - x) \end{bmatrix} \\
&\propto \begin{bmatrix} s_2(x) & -s_1(x) \\ -s_1(x) & \sum_{i=1}^n w_i \end{bmatrix} \begin{bmatrix} w_1 & \dots & w_n \\ w_1(x_1 - x) & \dots & w_n(x_n - x) \end{bmatrix} \\
&= \begin{bmatrix} w_1(s_2(x) - (x_1 - x)s_1(x)) & \dots & w_n\,(s_2(x) - (x_n - x)s_1(x)) \\ w_1((x_1 - x)\sum_{i=1}^n w_i - s_1(x)) & \dots & w_n((x_n - x)\sum_{i=1}^n w_i - s_1(x)) \end{bmatrix}.
\end{aligned}$$

From this we conclude that $\hat{f}(x)$ is a linear smoother with a weight on each $y_i$ proportional to $b_{x,i} = w_i(s_2(x) - (x_i - x)s_1(x))$.

(C)

(D) With $H$ a smoothing matrix (or "hat matrix"), let $r = y - Hy$ be the vector of residuals. If the random vector $x$ with mean vector $\mu$ and covariance matrix $\Sigma$, then $E(x^T Q x) = \text{tr}(Q\Sigma) + \mu^T Q \mu$. By assumption, $E(\mathbf{y}) = f(\mathbf{x})$ and $\text{cov}(\mathbf{y}) = \sigma^2 I$ Then,

$$\begin{aligned}
E\left(\|r^2\|_2^2\right) &= E\left((\mathbf{y} - H\mathbf{y})^T(\mathbf{y} - H\mathbf{y})\right) \\
&= E\left(\mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T H\mathbf{y} + \mathbf{y}^T H^T H\mathbf{y}\right) \\
&= E\left(\mathbf{y}^T\mathbf{y}\right) - 2E\left(\mathbf{y}^T H\mathbf{y}\right) + E\left(\mathbf{y}^T H^T H\mathbf{y}\right) \\
&= \left(\text{tr}\left[I\sigma^2 I\right] + f(\mathbf{x})^T f(\mathbf{x})\right) - 2\left(\text{tr}\left[H^T\sigma^2 I\right] + f(\mathbf{x})^T H^T f(\mathbf{x})\right) + \left(\text{tr}\left[H^T H\sigma^2 I\right] + f(\mathbf{x})^T H^T H f(\mathbf{x})\right) \\
&= \left(n\sigma^2 + f(\mathbf{x})^T f(\mathbf{x})\right) - 2\left(\sigma^2\text{tr}[H] + f(\mathbf{x})^T H^T f(\mathbf{x})\right) + \left(\sigma^2\text{tr}[H^T H] + f(\mathbf{x})^T H^T H f(\mathbf{x})\right) \\
&= \left(n - \text{tr}[H] + \text{tr}[H^T H]\right)\sigma^2 + \left(f(\mathbf{x})^T f(\mathbf{x}) - 2f(\mathbf{x})^T H^T f(\mathbf{x}) + f(\mathbf{x})^T H^T H f(\mathbf{x})\right) \\
&= \left(n - \text{tr}[H] + \text{tr}[H^T H]\right)\sigma^2 + (f(\mathbf{x}) - Hf(\mathbf{x}))^T\left(f(\mathbf{x}) - Hf(\mathbf{x})\right),
\end{aligned}$$

so the estimator

$$\hat{\sigma}^2 = \frac{\|r^2\|_2^2}{n - \text{tr}[H] + \text{tr}[H^T H]}$$

will be nearly unbiased in $\sigma^2$ when $f(\mathbf{x}) \approx Hf(\mathbf{x})$.

(E) See attached R code for implementation of the local polynomial regression, along with leave-one-out cross validation.

(F) Fitting this model with $D = 1$, the assumption of homoscedasticity (constance variance of residuals) is not met. The residuals fan out towards the lower end of the range for temperature. Taking a logarithmic transform of the response variable, daily gas bill, makes the residuals more uniform, although there are still several outliers in the residuals, now towards the higher ejnd of the range for temperature. See figures below.

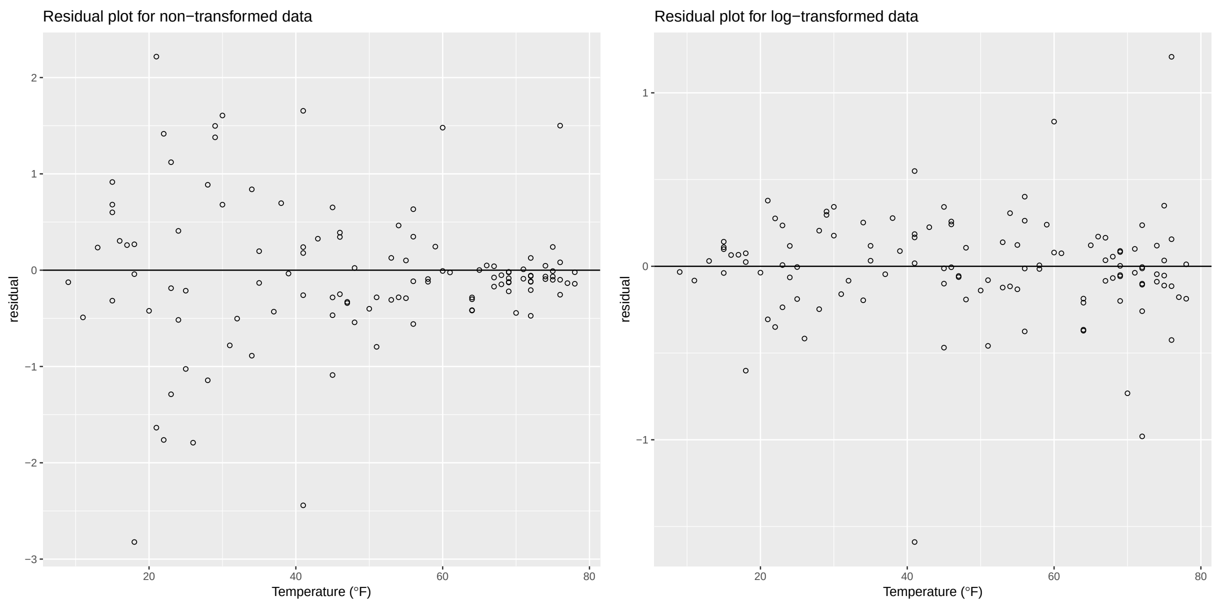


Figure 4: Residual plots for models fitted with non-transformed (left) and log-transformed (right) response variables

(G) The figure below shows the fitted model with 95% confidence bands (found with $\hat{y} \pm 1.96\hat{\sigma}$) and overlaid scatterplot of the data. The optimal bandwidth $h = 5.4168$ was chosen with leave-one-out cross validation. The fit is pretty good, with $R^2 = 0.88$, but there are four observations which fall outside the confidence band.

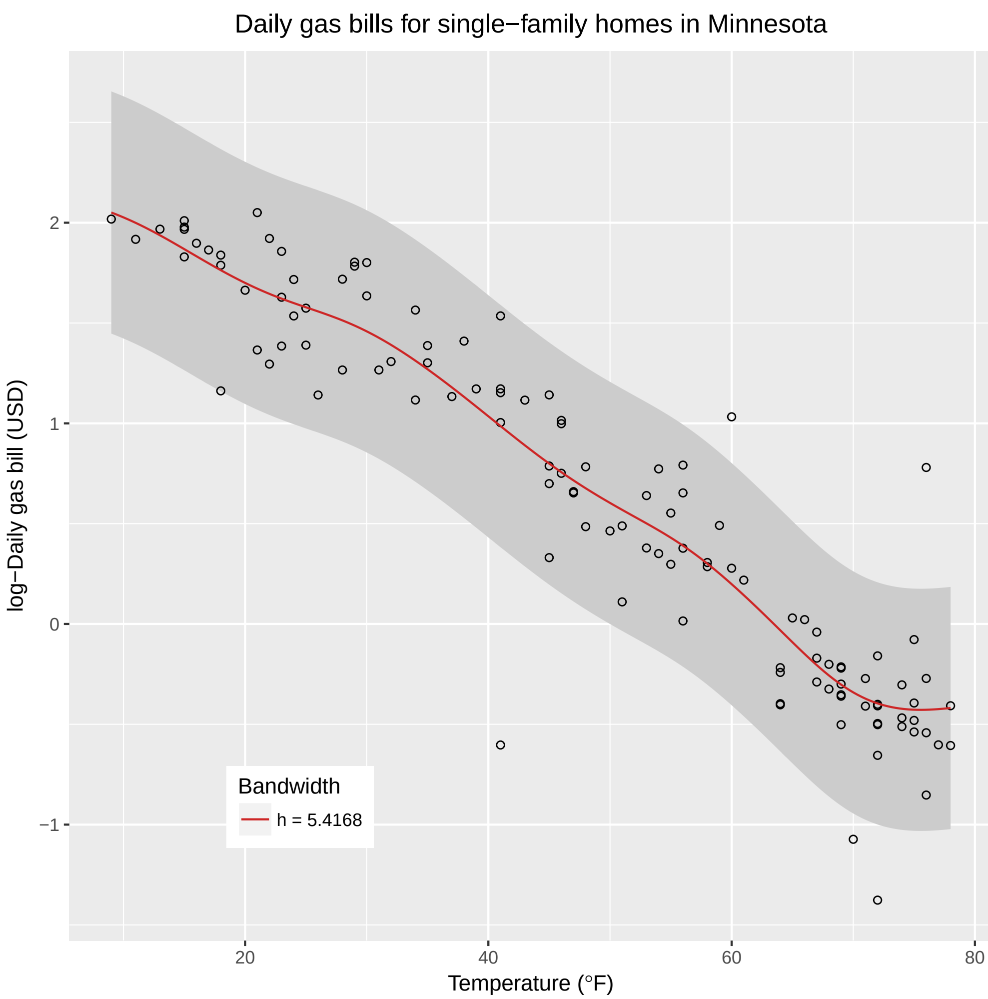


Figure 5: Local linear regression for Minnesota gas bill data

# Problem 5

## Gaussian processes

(A)

# Problem 6

## In nonparametric regression and spacial smoothing

(A)

R code for `myfuns03.R`

```
#############################################################
######### Created by Spencer Woody on 11 Feb 2017 #########
#############################################################

# =============================================================================
# Linear smoothing ============================================================
# =============================================================================

lin.smooth <- function(x.new, x, y, kern.fun, h) {
    # -------------------------------------------------------------------------
    # Linear smoother for some kernel function
    # -------------------------------------------------------------------------
    # INPUTS:
    # x.new -- a single new point for which to estimate f(x.new)
    # x -- a vector of covariates from previous observations
    # y -- a vector of responses from previous observations
    # kern.fun -- some kernel function (e.g. Gaussian)
    #             *** takes 2 arguments: distance (dist) and bandwidth (h)
    # h is the bandwidth for the kernel function
    # -------------------------------------------------------------------------
    # OUTPUT:
    # weights -- a vector of weights for a new observation
    # -------------------------------------------------------------------------

    weights <- kern.fun(dist = x - x.new, h = h) / h
    weights <- weights / sum(weights)

    fit <- crossprod(weights, y)

    return(fit)
}

kern.unif <- function(dist, h) {
    # -------------------------------------------------------------------------
    # Uniform kernel function
    # -------------------------------------------------------------------------
    # INPUTS:
    # dist
    # h is
    # Sigma is the covariance matrix
    # -------------------------------------------------------------------------
    # OUTPUT:
    # kern -- the value of the uniform kernel function
    # -------------------------------------------------------------------------

    kern <- ( abs(dist / h) <= 1) / 2

    return(kern)
}

kern.norm <- function(dist, h) {
    # -------------------------------------------------------------------------
```

```
    # Gaussian (normal) kernel function
    # ---------------------------------------------------------------------
    # INPUTS:
    # dist the distance
    # h is the bandwidth
    # Sigma is the covariance matrix
    # ---------------------------------------------------------------------
    # OUTPUT:
    # kern is the value of the Gaussian kernel function
    # ---------------------------------------------------------------------

    kern <- 1 / sqrt(2 * pi) * exp(-(dist / h)^2 / 2)

    return(kern)
}

make.noise <- function(x, f, res.dist, sd = NA, scale = NA, df = NA) {
    # ---------------------------------------------------------------------
    # Simulate noisy response from some non-linear function
    # ---------------------------------------------------------------------
    # INPUTS:
    # x - the predictor values
    # f - a function for the expected value, E(y | x) = f(x)
    # res.dist - a string for distribution of errors, either
    #             "normal" for normal errors, or
    #             "cauchy" for cauchy error
    # sd - the standard deviation for residuals (for normal errors only)
    # scale - the scale paramater for residuals (for cauchy errors only)
    # ---------------------------------------------------------------------
    # OUTPUT:
    # noise - the simulated noisy responses
    # ---------------------------------------------------------------------

    if (res.dist == "normal" & !is.na(sd)) {
        noisy <- f(x) + rnorm(n = length(x), mean = 0, sd = sd)
    } else if (res.dist == "cauchy" & !is.na(scale)) {
        noisy <- f(x) + rcauchy(n = length(x), location = 0, scale = scale)
    } else {
        stop(paste("Must give res.dist argument as ",
              "either \"normal\" or \"cauchy\", AND ",
              "also specify sd (for normal) or ",
              "scale (for cauchy)", sep = ""))
    }

    return(noisy)
}

# ===========================================================================
# Cross-validation ==========================================================
# ===========================================================================

cv <- function(x.tr, y.tr, x.te, y.te, KERN.FUN, h) {
    # ---------------------------------------------------------------------
```

```r
    # Give cross validation mean squared prediction error
    # ---------------------------------------------------------------------
    # INPUTS:
    # x.tr - vector of predictors in * training * set
    # y.tr - vector of responses in * training * set
    # x.te - vector of predictors in * testing * set
    # y.te - vector of responses in * testing * set
    # KERN.FUN - some kernel function (e.g. Gaussian)
    #            *** takes 2 arguments: distance (dist) and bandwidth (h)
    # h - vector or bandwidths
    # ---------------------------------------------------------------------
    # OUTPUT:
    # mse - mean square predictive error
    # ---------------------------------------------------------------------

    numbands <- length(h)
    mse <- rep(NA, numbands)

    for (i in 1:numbands) {
        y.pr <- sapply(
            x.te,
            lin.smooth,
            x = x.tr,
            y = y.tr,
            kern.fun = KERN.FUN,
            h = h[i]
            )

        mse[i] <- mean((y.te - y.pr)^2)
    }

    return(mse)
}

# ===========================================================================
# Local polynomial regression ===============================================
# ===========================================================================

loc.lin <- function(x.new, x.vec, y.vec, h) {
    # ---------------------------------------------------------------------
    # Give local linear estimator at some new point with normal kernel
    # ---------------------------------------------------------------------
    # INPUTS:
    # x.new is some new point on x
    # x.vec - vector of x in sample
    # y.vec - vector of y in sample
    # h - bandwidth
    # ---------------------------------------------------------------------
    # OUTPUT:
    # fit - the estimated value of f at x using local linear estimator
    # ---------------------------------------------------------------------

    kern.x <- kern.norm(x.new - x.vec, h)
```

```
    s1 <- sum(kern.x * (x - x.new))
    s2 <- sum(kern.x * (x - x.new)^2)
    # w.x <- kern.x * (s2 * x.new - s1 * (x - x.new))
    w.x <- kern.x * (s2 - s1 * (x - x.new))

    fit <- crossprod(w.x, y.vec) / sum(w.x)

    return(fit)
}

loc.pol <- function(x.new, x.vec, y.vec, D, h, give.mat = FALSE) {
    # ----------------------------------------------------------------------
    # Local polynomial reg. estimator at some new point with normal kernel
    # ----------------------------------------------------------------------
    # INPUTS:
    # x.new is some new point on x
    # x.vec - vector of x in sample
    # y.vec - vector of y in sample
    # D - dimension of the polynomial estimator (e.g., D = 1 means linear)
    # h - bandwidth for normal kernel
    # ----------------------------------------------------------------------
    # OUTPUT (list):
    # fit - the estimated value of f at x using local linear estimator
    # hatmat.vec - normalized weights, to be used in creating a hat matrix
    # ----------------------------------------------------------------------

    # Number of observations
    N <- length(x.vec)

    # Vector of weights from Gaussian kernel
    w <- kern.norm(x.vec - x.new, h) / h
    w <- w / sum(w)

    # Create (non-normalized weights)
    if (D == 0) { # case of local constant estimator

        b <- matrix(w, nrow = 1)

    } else { # case of local polynomial estimator

        # Create R matrix
        R.x <- matrix(nrow = N, ncol = (D + 1))
        R.x[, 1] <- rep(1, N)

        for (j in 2:(D+1)) {
            R.x[, j] <- (x.vec - x.new)^(j - 1)
        }

        W.diag <- diag(w)

        RxTW <- crossprod(R.x, W.diag)
```

```
        B <- solve(RxTW %*% R.x) %*% RxTW
        b <- matrix(B[1, ], nrow = 1)
    }

    fit <- tcrossprod( (b / sum(b)) , y.vec)

    # Should the hat matrix vector be given?
    if (give.mat) {
        output <- list("fit" = fit, "hatmat.vec" = as.vector(b / sum(b)))
    } else {
        output <- fit
    }

    return(output)
}

loc.pol.hatmat <- function(x, y, D, h) {
    # ----------------------------------------------------------------------
    # Create a hat matrix from local polynomial estimator
    # ----------------------------------------------------------------------
    # INPUTS:
    # x - vector of x in sample
    # y - vector of y in sample
    # D - dimension of the polynomial estimator (e.g., D = 1 means linear)
    # h - bandwidth for normal kernel
    # ----------------------------------------------------------------------
    # OUTPUT (list):
    # hatmat - hat matrix from local polynomial estimator
    # ----------------------------------------------------------------------

    N <- length(x)

    hatmat <- matrix(nrow = N, ncol = N)

    for (i in 1:N) {
        hatmat[i, ] <- loc.pol(x[i], x, y, D, h, give.mat = TRUE)$hatmat.vec
    }

    return(hatmat)
}

loocv <- function(y, H) {
    # ----------------------------------------------------------------------
    # Generic leave-one-out cross validation
    # ----------------------------------------------------------------------
    # INPUTS:
    # y - the response vector
    # H - the "hat" matrix
    # ----------------------------------------------------------------------
    # OUTPUT:
    # loocv - leave-one-out cross validation error
    # ----------------------------------------------------------------------
```

```
    y.hat <- H %*% y

    loocv <- sum( ( {y - y.hat} / {1 - diag(H)} )^2 )

    return(loocv)
}

# ===========================================================================
# Gaussian process ==========================================================
# ===========================================================================

my.mvn <- function(n, mu, Sigma) {
    #  Simulate n draws from MVN(mu, Sigma)
    #
    #  Note: this function assumes that X already has an intercept term
    # (or doesn't, if we want to force OLS through the origin)
    #
    # INPUTS:
    # n is the number of draws
    # mu is the mean vector
    # Sigma is the covariance matrix
    #
    # OUTPUT:
    # x is matrix of n draws from MVN(mu, Sigma) [with n rows, p columns]
    #

    # dimension of MVN
    p <- length(mu)

    # Check if inputs are valid (dimensions match, Sigma is square and p.s.d.)
    cond<- (ncol(Sigma) != p) |
           (nrow(Sigma) != p) |
           (max(eigen(Sigma)$values) <= 0)

    if (cond) {
        return("Try again...")
    }

    # Generate n*p univariate standard normal variables
    z     <- matrix(rnorm(n*p), nrow = p)

    # Create a matrix containing copies of mu
    mumat <- matrix(rep(mu, n), nrow = p)

    # Decompose Sigma into Sigma = L %*% Lt
    Lt <- chol(Sigma)

    # Generate sample with affine transformation of z
    x <- crossprod(Lt, z) + mumat

    return(t(x))
}
```

```
ell2 <- function(x) {
    # Compute the ell2 norm of x, a vector in Euclidean space

    return(sqrt(sum(x^2)))
}

C.SE <- function(x.i, x.j, params = NA) {
    # ------------------------------------------------------------------
    #  Compute the (i, j) element of a squared exp. covariance matrix
    #
    # ------------------------------------------------------------------
    # INPUTS:
    # x.i and x.j are two vectors in same space (need not be [0, 1])
    # params should be a vector of three hyperparameters
    #      1) b
    #      2) tau1.sq
    #      3) tau2.sq
    #
    # ------------------------------------------------------------------
    # OUTPUT:
    # c.se is the value of the Matern-5/2 covariance matrix for x.i and x.j
    # ------------------------------------------------------------------

    if (prod(is.na(params))) {
        return("Must have three valid parameters.")
    }

    if (length(params) != 3) {
        return("Must have three valid parameters.")
    }

    b       <- params[1]
    tau1.sq <- params[2]
    tau2.sq <- params[3]


    b       <- params[1]
    tau1.sq <- params[2]
    tau2.sq <- params[3]

    # Euclidean distance between x.i and x.j
    d <- ell2(x.i - x.j)

    c.se <- tau1.sq * exp(-0.5 * (d / b)^2) + tau2.sq * (x.i == x.j)

    return(c.se)
}

C.M52 <- function(x.i, x.j, params = NA) {
    # ------------------------------------------------------------------
    #  Compute the (i, j) element of a Matern-5/2 covariance matrix
    #
    # ------------------------------------------------------------------
```

```
    # INPUTS:
    # x.i and x.j are two vectors in same space (need not be [0, 1])
    # params should be a vector of three hyperparameters
    #       1) b
    #       2) tau1.sq
    #       3) tau2.sq
    #
    # ----------------------------------------------------------------------
    # OUTPUT:
    # c.m52 is the value of the Matern-5/2 covariance matrix for x.i and x.j
    # ----------------------------------------------------------------------

    if (prod(is.na(params))) {
        return("Must have three valid parameters.")
    }

    if (length(params) != 3) {
        return("Must have three valid parameters.")
    }

    b       <- params[1]
    tau1.sq <- params[2]
    tau2.sq <- params[3]

    # Euclidean distance between x.i and x.j
    d <- ell2(x.i - x.j)

    c.m52 <- tau1.sq * ( 1 + (5^0.5 * d / b) + (5 / 3 * (d / b)^2) ) *
             exp(-5^0.5 * d / b)  + tau2.sq * (x.i == x.j)

    return(c.m52)
}

make.covmat <- function(x, cov.fun, params = NA) {
    # ----------------------------------------------------------------------
    #  Compute the covariance matrix for a GP, given some cov. function
    #
    # ----------------------------------------------------------------------
    # INPUTS:
    # x is a vector of N values in [0, 1]
    # params should be a vector of three hyperparameters
    #       1) b
    #       2) tau1.sq
    #       3) tau2.sq
    #
    # ----------------------------------------------------------------------
    # OUTPUT:
    # covmat is the covariance matrix of GP
    # ----------------------------------------------------------------------

    if (prod(is.na(params))) {
        return("Must have three valid parameters.")
    }
```

```
    if (length(params) != 3) {
        return("Must have three valid parameters.")
    }

    N <- length(x)

    covmat <- matrix(nrow = N, ncol = N)

    for (j in 1:N) {
        for (i in j:N) {
            covmat[i, j] <- cov.fun(x[i], x[j], params = params)
            covmat[j, i] <- covmat[i, j]
        }
    }

    return(covmat)
}
```

R code for exercises03.R

```
##########################################################
######### Created by Spencer Woody on 11 Feb 2017 #########
##########################################################

library(ggplot2)
library(reshape2)
library(gridExtra)
library(wesanderson) # nice palettes

# Prep color palette
# pal <- wes_palette("Zissou", 5)
# col1 <- pal[5]
# col2 <- pal[4]
# col3 <- pal[1]

col1 <- "red"
col2 <- "orange"
col3 <- "blue"

source("myfuns03.R")

# ==========================================================================
# Linear smoothers (part one) ==============================================
# ==========================================================================

# nonlinear function f(x)
f1 <- function(x){
    return(x * (x - 4) * (x + 4))
}

# Predictor vector
x1 <- seq(-5, 5, length.out = 40)

# Create sequence along x-space
x.seq <- seq(min(x1), max(x1), length.out = 200)

# Response vector
y1 <- make.noise(x1, f1, "normal", sd = 15)

# Bin width
h1 <- 0.75

# Gaussian kernel smoothing
y.norm <- sapply(
    x.seq,
    lin.smooth,
    x = x1,
    y = y1,
    kern.fun = kern.norm,
    h = h1
    )
```

```
# Uniform kernel smoothing
y.unif <- sapply(
    x.seq,
    lin.smooth,
    x = x1,
    y = y1,
    kern.fun = kern.unif,
    h = h1
    )

# Make a nice plot
h <- qplot(x.seq, geom = "blank") +
xlab("x") +
ylab("y") +
ggtitle(sprintf("Smoothing of cubic function")) +
geom_point(aes(x = x1, y = y1), pch = 1) +
stat_function(fun = f1, col = col1, linetype = "dashed") +
geom_line(aes(y = y.norm, colour = "Gaussian kernel")) +
geom_line(aes(y = y.unif, colour = "Uniform kernel")) +
scale_colour_manual(name = "Smoother", values = c(col3, col2)) +
theme(legend.position = c(0.25, 0.15),
    text = element_text(family="Helvetica"))

h

pdf("firstexample.pdf")
h
dev.off()



# ==========================================================================
# Linear smoothers (cross validation) ==== make big heat map ================
# ==========================================================================

# Sample size
N = 500

# Set limits of x-space
xlo = 0
xhi = 1

# Number of bins for cross-validation
numbins <- 5

# Vector for bandwidths to search over
h.vec <- seq(0.001, 0.125, length.out = 100)

# Vector for standard deviations
s.vec <- seq(0.001, 0.5, length.out = 100)
s.vec <- rev(s.vec)

# Vector for periods
```

```
p.vec <- seq(0.1, 1, length.out = 100)

# Matrix for optimal bandwidth values
opt.h <- matrix(nrow = length(p.vec), ncol = length(s.vec))

for (i in 1:length(p.vec)) {
    # Select what the period is for this iteration
    p.i <- p.vec[i]

    # Create a new function with this current period
    mysin.i <- function(x) {
        return(sin(x * 2*pi / p.i))
    }

    for (j in 1:length(s.vec)) {
        # Select what the residual sd is for this iteration
        s.j <- s.vec[j]

        # Generate data
        x.ij <- (xhi - xlo) * runif(N) + xlo
        y.ij <- make.noise(x.ij, mysin.i, "normal", sd = s.j)

        # Prepare mse matrix for this current iteration
        mse.ij <- matrix(nrow = numbins, ncol = length(h.vec))

        # Create random partition into five bins
        jumble <- sample(1:N, N, replace = F)
        bin.indices <- split(jumble, cut(1:N, numbins))

        for (bin in 1:numbins) {
            my.indices <- bin.indices[[bin]]
            x.tr <- x.ij[-(my.indices)]
            y.tr <- y.ij[-(my.indices)]
            x.te <- x.ij[my.indices]
            y.te <- y.ij[my.indices]
            mse.ij[bin, ] <- cv(x.tr, y.tr, x.te, y.te, kern.norm, h.vec)
        }

        # Average out MSE over all bins
        mse.vec <- colMeans(mse.ij)

        # Choose bandwidth with lowest average MSE
        opt.h[i, j] <- h.vec[which.min(mse.vec)]

    }
    print(i)
}

# Check to make sure optimal bandwidths look OK
opt.h / max(opt.h)

# Plot these things
ohm <- melt(opt.h)
```

```r
w <- ggplot(ohm, aes(rev(Var1), Var2, z = value)) +
ggtitle("Picking Optimal Bandwidth for Gaussian Kernel Smoothing (5-fold CV)") +
xlab("Decreasing standard deviation (0.5:0.001) (less noisy)") +
ylab("Decreasing period (1:0.1) (more wiggly)") +
geom_tile(aes(fill = value)) +
scale_fill_distiller("Bandwidth", palette = "Spectral")
w

# Save this to PDF
pdf("img/opth.pdf", width = 7, height = 6)
w
dev.off()


# =============================================================================
# Linear smoothers (cross validation) ==== make 2 x 2 plot ==================
# =============================================================================

# Sample size
N = 500

# Set limits of x-space
xlo = 0
xhi = 1

# Number of bins for cross-validation
numbins <- 5

# Vector for bandwidths to search over
h.vec <- seq(0.001, 0.125, length.out = 100)

# Vector for standard deviations
# s.vec <- seq(0.001, 0.5, length.out = 64)
s.vec <- c(0.1, 0.5)

# Vector for periods
# p.vec <- seq(0.1, 1, length.out = 64)
p.vec <- c(0.125, 1)

# Matrix for optimal bandwidth values
opt.h <- matrix(nrow = length(p.vec), ncol = length(s.vec))

# Matrix for random x-values and y-values
x.mat <- matrix(nrow = nrow(opt.h) * ncol(opt.h), ncol = N)
y.mat <- matrix(nrow = nrow(opt.h) * ncol(opt.h), ncol = N)

# Matrix for values of f at x
x.seq <- seq(xlo, xhi, length.out = 200)
fx.mat <- matrix(nrow = nrow(opt.h) * ncol(opt.h), ncol = length(x.seq))
smooth <- matrix(nrow = nrow(opt.h) * ncol(opt.h), ncol = length(x.seq))

count <- 1
```

```
for (i in 1:length(p.vec)) {
    # Select what the period is for this iteration
    p.i <- p.vec[i]

    # Create a new function with this current period
    mysin.i <- function(x) {
        return(sin(x * 2*pi / p.i))
    }

    for (j in 1:length(s.vec)) {
        # Select what the residual sd is for this iteration
        s.j <- s.vec[j]

        x.ij <- (xhi - xlo) * runif(N) + xlo
        y.ij <- make.noise(x.ij, mysin.i, "normal", sd = s.j)

        fx.mat[count, ] <- mysin.i(x.seq)

        x.mat[count, ] <- x.ij
        y.mat[count, ] <- y.ij

        mse.ij <- matrix(nrow = numbins, ncol = length(h.vec))

        jumble <- sample(1:N, N, replace = F)
        bin.indices <- split(jumble, cut(1:N, numbins))

        for (bin in 1:numbins) {
            my.indices <- bin.indices[[bin]]
            x.tr <- x.ij[-(my.indices)]
            y.tr <- y.ij[-(my.indices)]
            x.te <- x.ij[my.indices]
            y.te <- y.ij[my.indices]
            mse.ij[bin, ] <- cv(x.tr, y.tr, x.te, y.te, kern.norm, h.vec)
        }

        mse.vec <- colMeans(mse.ij)

        # Choose first 4/5 of x's and y's to be training, the rest of testing
        # x.tr <- x.ij[1:round(N*4/5)]
        # y.tr <- y.ij[1:round(N*4/5)]
        # x.te <- x.ij[-(1:round(N*4/5))]
        # y.te <- y.ij[-(1:round(N*4/5))]

        # mse.vec <- cv(x.tr, y.tr, x.te, y.te, kern.norm, h.vec)

        opt.h[i, j] <- h.vec[which.min(mse.vec)]

        smooth[count, ] <- y.norm <- sapply(
                            x.seq,
                            lin.smooth,
                            x = x.ij,
                            y = y.ij,
```

```
                              kern.fun = kern.norm,
                              h = opt.h[i, j]
                              )


        count <- count + 1
    }
    print(i)
}

y.min <- min(y.mat)
y.max <- max(y.mat)

ess <- qplot(x.seq, geom = "blank") +
xlab("x") +
ylab("y") +
ylim(y.min, y.max) +
ggtitle(sprintf("sd = %5.3f, T = %5.3f, h = %5.5f", s.vec[1], p.vec[1], opt.h[1, 1])) +
geom_point(aes(x = x.mat[1, ], y = y.mat[1, ]), pch = 1) +
geom_line(aes(y = fx.mat[1, ]), col = "red", linetype = "dashed") +
geom_line(aes(y = smooth[1, ], colour = "Gaussian kernel")) +
scale_colour_manual(name = "Smoother", values = "blue") +
theme(legend.position = c(0.25, 0.15), text = element_text(family="Helvetica"))

tee <- qplot(x.seq, geom = "blank") +
xlab("x") +
ylab("y") +
ylim(y.min, y.max) +
ggtitle(sprintf("sd = %5.3f, T = %5.3f, h = %5.5f", s.vec[2], p.vec[1], opt.h[1, 2])) +
geom_point(aes(x = x.mat[2, ], y = y.mat[2, ]), pch = 1) +
geom_line(aes(y = fx.mat[2, ]), col = "red", linetype = "dashed") +
geom_line(aes(y = smooth[2, ], colour = "Gaussian kernel")) +
scale_colour_manual(name = "Smoother", values = "blue") +
theme(legend.position = c(0.25, 0.15), text = element_text(family="Helvetica"))

you <- qplot(x.seq, geom = "blank") +
xlab("x") +
ylab("y") +
ylim(y.min, y.max) +
ggtitle(sprintf("sd = %5.3f, T = %5.3f, h = %5.5f", s.vec[1], p.vec[2], opt.h[2, 1])) +
geom_point(aes(x = x.mat[3, ], y = y.mat[3, ]), pch = 1) +
geom_line(aes(y = fx.mat[3, ]), col = "red", linetype = "dashed") +
geom_line(aes(y = smooth[3, ], colour = "Gaussian kernel")) +
scale_colour_manual(name = "Smoother", values = "blue") +
theme(legend.position = c(0.25, 0.15), text = element_text(family="Helvetica"))


vee <- qplot(x.seq, geom = "blank") +
xlab("x") +
ylab("y") +
ylim(y.min, y.max) +
ggtitle(sprintf("sd = %5.3f, T = %5.3f, h = %5.5f", s.vec[2], p.vec[2], opt.h[2, 2])) +
geom_point(aes(x = x.mat[4, ], y = y.mat[4, ]), pch = 1) +
geom_line(aes(y = fx.mat[4, ]), col = "red", linetype = "dashed") +
```

```
geom_line(aes(y = smooth[4, ], colour = "Gaussian kernel")) +
scale_colour_manual(name = "Smoother", values = "blue") +
theme(legend.position = c(0.25, 0.15), text = element_text(family="Helvetica"))


pdf("img/2x2.pdf")
grid.arrange(tee, ess, vee, you)
dev.off()

# ============================================================================
# Local polynomial regression ================================================
# ============================================================================

# ----------------------------------------------------------------------------
# Read in the data -----------------------------------------------------------
# ----------------------------------------------------------------------------

utilities <- read.csv("utilities.csv", header = T)

x <- utilities$temp
y <- log(utilities$gasbill / utilities$billingdays)


# ----------------------------------------------------------------------------
# Cross validation -----------------------------------------------------------
# ----------------------------------------------------------------------------

h.vec <- seq(1, 20, length.out = 500)
num.h <- length(h.vec)

hatmat.list <- list()
loocv.vec <- rep(NA, num.h)

for (i in 1:num.h) {
    hatmat.i <- loc.pol.hatmat(x, y, D = 1, h = h.vec[i])
    hatmat.list[[i]] <- hatmat.i

    loocv.vec[i] <- loocv(y, hatmat.i)

    if ((i %% 20) == 0) {
        print(sprintf("Iteration %i out of %i...", i, num.h))
    }
}

plot(h.vec, loocv.vec, type = "l", xlab = "Bandwidth", ylab = "LOOCV")

h.opt <- h.vec[which.min(loocv.vec)]

# ----------------------------------------------------------------------------
# Make a plot ----------------------------------------------------------------
# ----------------------------------------------------------------------------

# h.opt <- 4.9
```

```
x.seq <- seq(min(x), max(x), length.out = 200)

y.smooth <- sapply(
    x.seq,
    loc.pol,
    x.vec = x,
    y.vec = y,
    D = 1,
    h = h.opt
    )


# Fitted y values
Hatmat <- loc.pol.hatmat(x, y, D = 1, h = h.opt)
y.hat <- Hatmat %*% y

# R—squared
r.sq <- 1 - sum((y - y.hat)^2) / sum((y - mean(y))^2)

# Estimated variance
var.est <- sum((y - y.hat)^2) /
(length(x) - sum(diag(Hatmat)) + sum(diag(crossprod(Hatmat))))

y.lo <- y.smooth - 1.96 * sqrt(var.est)
y.hi <- y.smooth + 1.96 * sqrt(var.est)


resplot <- qplot(x, y - y.hat, geom = "blank") +
geom_point(pch = 1) +
geom_hline(yintercept = 0) +
xlab(expression(paste("Temperature (",degree,"F)"))) +
ylab("residual") +
ggtitle("Residual plot for log-transformed data")

pdf("img/resplot2.pdf")
resplot
dev.off()

q <- qplot(x.seq, geom = "blank") +
xlab(expression(paste("Temperature (",degree,"F)")))  +
ylab("log-Daily gas bill (USD)") +
labs(title = "Daily gas bills for single-family homes in Minnesota") +
geom_ribbon(aes(ymin = y.lo, ymax = y.hi), fill = "grey80") +
geom_point(aes(x = x, y = y), pch = 1) +
geom_line(aes(y = y.smooth, colour = sprintf("h = %5.4f", h.opt)))  +
scale_colour_manual(name = "Bandwidth", values = "firebrick3") +
theme(plot.title = element_text(hjust = 0.5),
text = element_text(family = "Helvetica"),
legend.position = c(0.25, 0.15))


pdf("img/tempplot.pdf")
```

```
q
dev.off()

# R-squared
# 1 - sum((loc.pol.hatmat(x, y, 1, 5) %*% y - y)^2) / sum((y - mean(y))^2)

# ============================================================================
# Gaussian process ===========================================================
# ============================================================================

x.seq <- seq(0, 1, length.out = 100)

b <- 1
tau1.sq <- 1e-6
tau2.sq <- 1e-5

myparams <- c(b, tau1.sq, tau2.sq)

xCM52 <- make.covmat(x.seq, C.M52, params = myparams)
xSE <- make.covmat(x.seq, C.SE, params = myparams)


# ============================================================================
# Extra code for CV plotting.... ============================================
# ============================================================================

# Cross-validation with error bars

RSS.vec <- apply(RSS.mat, 2, mean)
RSS.SE <- apply(RSS.mat, 2, sd) / sqrt(numbins)

lower = RSS.vec - RSS.SE
upper = RSS.vec + RSS.SE

# Make a plot!
pdf("perror.pdf", width = 12 / 1.25, height = 8 / 1.25)
h <- qplot(log(fit1$lambda), RSS.vec, geom = "path")
h + xlab(expression(paste("Penalization term, log(", lambda, ")"))) +
ylab("Expected prediction error") +
labs(title = sprintf("%i-fold Cross-validation, Mallow's CP, and In-sample MSE", numbins)) +
geom_ribbon(aes(ymin = lower, ymax = upper), fill = "grey80") +
geom_line(aes(y = RSS.vec, colour = "CV"), show.legend = TRUE) +
geom_line(aes(y = MSE.lasso, colour = "In-sample MSE"), show.legend = TRUE) +
geom_line(aes(y = MCP, colour = "CP"), show.legend = TRUE) +
geom_vline(xintercept = log(minlambdaCV), linetype = 3, col = "black", size = 0.75, show.legend = TRU
geom_vline(xintercept = log(minlambdaMCP), linetype = 3, col = "red", size = 0.75, show.legend = TRUE
scale_colour_manual(name = "", values = c("CV" = "black", "In-sample MSE" = "blue", "CP" = "red"))
dev.off()


numbins <- 10
jumble <- sample(1:N2, N2, replace = F)
bin.indices <- split(jumble, cut(1:N2, numbins))
```