We have that the estimated density from the kernel density estimator is given by

$$\hat{f}_\theta(x) = \frac{1}{n}\sum_{i=1}^{n} K_\theta(x - x'_i) \tag{1}$$

where $K_h$ is the kernel function, and $\theta$ are the parameters, and $x'_i$ is the $i$th training point. In our case $x'_i$ is a $m$ dimensional vector representing the daily change for each of the $m$ stocks in our dataset. We assume that that each kernel function is normalized $\int_{\mathbb{R}^d} K_h(x)dx = 1$

# Optimizing the Kernel Parameters

We would want to minimize the weighted log likelihood function, ie we would want to minimize:

$$\mathcal{L}(x'_1, \ldots, x'_k) = -\sum_{i=1}^{k} w_i \log(\hat{f}_\theta(x'_i)) \tag{2}$$

Where are $w_i$ are weights of the $i$th training point, we will experiments with different weighting mechanisms later on.

## Gaussian Kernel

Let us consider the case of a gaussian kernel

$$K_\Sigma(x) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\left(-\frac{1}{2}x^T\Sigma^{-1}x\right)$$

Where $\Sigma$ is the covariance matrix for the kernel. Because the covariance matrix is symmetric positive semidefinite we can express $\Sigma$ as $\Sigma = LL^T$ where $L$ is a lower triangular matrix. We have that the derivative of the weighted log likelihood function is given by:

$$\frac{\partial}{\partial L^T}\mathcal{L}(x'_1, \ldots, x'_k) = -\sum_{i=1}^{k} w_i \frac{1}{\hat{f}_\theta(x'_i)} \frac{\partial \hat{f}_\theta(x'_i)}{\partial L^T}$$

We have that

$$\frac{\partial|\Sigma|}{\partial L^T} = \frac{\partial|LL^T|}{\partial L^T}$$
$$= 2|\Sigma|L^{-1}$$

We also have that: . We assume in our case that the kernel is scaled such that $\int_{\mathbb{R}^d} K_h(x)dx = 1$. What we want to optimize is the sharpe ratio, which is given by:

$$\frac{\mathbb{E}[R_p]}{\sigma_p} \tag{3}$$

If we have that the weights for each of the 10 stocks is given by a vector $w$, then we can write the expected return as:

$$\mathbb{E}[R_p] = \mathbb{E}[w^T x] \tag{4}$$

Where $x$ is the vector of daily returns for each of the 10 stocks. We have that this is given by:

$$\mathbb{E}[w^T x] = \int_{\mathbb{R}^{10}} w^T x \hat{f}_h(x)dx \tag{5}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}^{10}} w^T x K_h(x - x_i')dx \tag{6}$$

Now we can use the fact that the kernel function is symmetric to get:

$$\int_{\mathbb{R}^{10}} x K_h(x - x_i')dx = \int_{\mathbb{R}^{10}} (x_i - u)K_h(u)du$$

Where $u = x - x_i'$, because $K_h(u)$ is symmetric (even) about $u = 0$. we have that $\int_{\mathbb{R}^{10}} uK_h(u)du = 0$, thus we have that

$$\int_{\mathbb{R}^{10}} x K_h(x - x_i')dx = \int_{\mathbb{R}^{10}} x_i K_h(u)du \tag{7}$$
$$= x_i \tag{8}$$

Therefore we have that:

$$\mathbb{E}[w^T x] = \frac{1}{n}\sum_{i=1}^{n} w^T x_i' \tag{9}$$

Now we need to find the variance of the portfolio, which is given by:

$$\sigma_p^2 = \mathbb{E}[(w^T x)^2] - \mathbb{E}[w^T x]^2 \tag{10}$$

From equation (9) we have that, $\mathbb{E}[w^T x]^2 = \left(\frac{1}{n}\sum_{i=1}^n w^T x_i'\right)^2$. We have that $\mathbb{E}[(w^T x)^2]$ is given by:

$$\begin{aligned}
\mathbb{E}[(w^T x)^2] &= \int_{\mathbb{R}^{10}} (w^T x)^2 \hat{f}_h(x) dx && (11)\\
&= \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^{10}} (w^T x)^2 K_h(x - x_i') dx && (12)
\end{aligned}$$

We have that because $w^T x$ is a scalar, we have that $(w^T x)^T = x^T w = w^T x$, therefore we have that:

$$\begin{aligned}
\mathbb{E}[(w^T x)^2] &= \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^{10}} w^T x x^T w K_h(x - x_i') dx && (13)\\
&= \frac{1}{n}\sum_{i=1}^n w^T \left(\int_{\mathbb{R}^{10}} x x^T K_h(x - x_i') dx\right) w && (14)\\
&= w^T \left(\frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^{10}} x x^T K_h(x - x_i') dx\right) w && (15)
\end{aligned}$$

Now let us limit our consideration to a gaussian kernel, ie

$$K_h(x - x_i') = \frac{1}{(2\pi)^{d/2} h} \exp\left(-\frac{1}{2}\frac{(x - x_i')^T (x - x_i')}{h^2}\right) \tag{16}$$

We can see that this is equivalent to a multivariate gaussian with $\mu = x_i'$ and $\Sigma = \frac{1}{h^2} I$. Therefore we have that $\int_{\mathbb{R}^{10}} x x^T K_h(x - x_i')$ is effectively the second moment of this multivariate gaussian, which is given by:

$$\int_{\mathbb{R}^{10}} x x^T K_h(x - x_i') dx = x_i x_i^T + \frac{1}{h^2} I \tag{17}$$

Therefore we have that equation (15) becomes

$$\mathbb{E}[(w^T x)^2] = w^T \left( \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T + \frac{1}{h^2} I \right) w \tag{18}$$

$$= w^T \left( \frac{1}{h^2} I + \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T \right) w \tag{19}$$

$$= \frac{1}{h^2} w^T w + w^T \left( \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T \right) w \tag{20}$$

Therefore we have that equation (10) becomes:

$$\sigma_p^2 = \frac{1}{h^2} w^T w + w^T \left( \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T \right) w - \left( \frac{1}{n} \sum_{i=1}^{n} w^T x_i' \right)^2 \tag{21}$$

And thus the sharpe ratio is given by:

$$\frac{\mathbb{E}[R_p]}{\sigma_p} = \frac{\frac{1}{n} \sum_{i=1}^{n} w^T x_i'}{\sqrt{\frac{1}{h^2} w^T w + w^T \left( \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T \right) w - \left( \frac{1}{n} \sum_{i=1}^{n} w^T x_i' \right)^2}} \tag{22}$$

Maximizing this is a fractional programming problem, which can be solved with known algorithms.