(Mercer's Condition) Necessary & Sufficient condition for a function $k(\underline{x}, \underline{x}')$ to be a valid kernel:

The Gram Matrix K should be PSD $\forall\ \underline{x}_i$ $\qquad K(i,j) = k(\underline{x}_i, \underline{x}_j)$

i.e., $\underline{x}_i^T K \underline{x}_i \geq 0$

Properties: given valid kernels $k_1(\underline{x}, \underline{x}')$ and $k_2(\underline{x}, \underline{x}')$, the following new kernels will also be valid kernels:-

1) $k(\underline{x}, \underline{x}') = c\, k_1(\underline{x}, \underline{x}')$, $c > 0$
2) $k(\underline{x}, \underline{x}') = f(\underline{x})\, k_1(\underline{x}, \underline{x}')\, f(\underline{x}')$
   $\longrightarrow f(\cdot)$ is a function
3) $k(\underline{x}, \underline{x}') = q\left(k_1(\underline{x}, \underline{x}')\right)$, $q(\cdot)$: polynomial with non-negative coefficients
4) $k(\underline{x}, \underline{x}') = \exp\left(k_1(\underline{x}, \underline{x}')\right)$
5) $k(\underline{x}, \underline{x}') = k_1(\underline{x}, \underline{x}') + k_2(\underline{x}, \underline{x}')$
6) $k(\underline{x}, \underline{x}') = k_1(\underline{x}, \underline{x}')\, k_2(\underline{x}, \underline{x}')$
7) $k(\underline{x}, \underline{x}') = k_3\left(\underline{\phi}(\underline{x}), \underline{\phi}(\underline{x}')\right)$, $\underline{\phi}(\underline{x}): \underline{x} \longrightarrow \mathbb{R}^M$
8) $k(\underline{x}, \underline{x}') = k_a(\underline{x}_a, \underline{x}_a') + k_b(\underline{x}_b, \underline{x}_b')$
   $\underline{x} = (\underline{x}_a, \underline{x}_b)$: $\underline{x}_a$ & $\underline{x}_b$ are not necessarily disjoint, $k_a(\cdot)$ and $k_b(\cdot)$ are valid kernel functions.
9) $k(\underline{x}, \underline{x}') = k_a(\underline{x}_a, \underline{x}_a')\, k_b(\underline{x}_b, \underline{x}_b')$

"Gaussian kernel"

Interesting Example: $k(\underline{x},\underline{x}') = \exp\left(-\frac{\|\underline{x}-\underline{x}'\|^2}{2\sigma^2}\right)$

$$\|\underline{x}-\underline{x}'\|^2 = (\underline{x}-\underline{x}')^T(\underline{x}-\underline{x}')$$
$$= (\underline{x}^T-\underline{x}'^T)(\underline{x}-\underline{x}')$$
$$= \underline{x}^T\underline{x} + \underline{x}'^T\underline{x}' - 2\underline{x}^T\underline{x}'$$

$$k(\underline{x},\underline{x}') = \exp\left(\frac{-\underline{x}^T\underline{x}}{2\sigma^2}\right)\cdot\exp\left(-\frac{\underline{x}'^T\underline{x}'}{2\sigma^2}\right)\cdot\exp\left(\frac{\underline{x}^T\underline{x}'}{\sigma^2}\right)$$

$$\exp\left(\frac{-\underline{x}^T\underline{x}}{2\sigma^2}\right)\underbrace{\exp\left(\frac{\underline{x}^T\underline{x}'}{\sigma^2}\right)}\exp\left(\frac{-\underline{x}'^T\underline{x}'}{2\sigma^2}\right)$$

$\underline{x}^T\underline{x}'$ is a kernel (linear kernel)

$\Rightarrow \left(\frac{1}{\sigma}\underline{x}^T\right)\left(\frac{1}{\sigma}\underline{x}'\right)$ is also a kernel

$\Rightarrow \frac{\underline{x}^T\underline{x}'}{\sigma^2}$ is a kernel

$\Rightarrow \underbrace{\exp\left(\frac{\underline{x}^T\underline{x}'}{\sigma^2}\right)}_{k_1(\underline{x},\underline{x}')}$ is also a kernel ($\because \exp(k_1(\underline{x},\underline{x}'))$ is a kernel)

$\Rightarrow f(\underline{x})\, k_1(\underline{x},\underline{x}')\cdot f(\underline{x}')$ is also a kernel

$f(\underline{x}) = \exp\left(\frac{-\underline{x}^T\underline{x}}{2\sigma^2}\right)$, $f(\underline{x}') = \exp\left(\frac{-\underline{x}'^T\underline{x}'}{2\sigma^2}\right)$

QED.