Elaboration on Kernel Machine Dot Product Idea

Kernel machines take the original feature vectors and transform them to a higher dimensional space in which they can be more easily discriminated. In the book's example, 2D points are transformed to 3D space. But, in general, we would not know what function F to use for this transformation. The dot product idea is that in the higher space, we use dot products and WE DON'T HAVE TO KNOW WHAT F IS, just what its dot product is.

In the book's example,
Each vector $\mathbf{x} = (x_1, x_2)$.
The function F transforms each vector to 3D space.

$F(\mathbf{x}) = (x_1^2, x_2^2, \sqrt{2}x_1x_2)$
$F(\mathbf{x_i}) \bullet F(\mathbf{x_j})$
= $(x_{1i}^2, x_{2i}^2, \sqrt{2}x_{1i}x_{2i}) \bullet (x_{1j}^2, x_{2j}^2, \sqrt{2}x_{1j}x_{2j})$
= $(x_{1i}x_{1j})^2 + 2x_{1i}x_{1j}x_{2i}x_{2j} + (x_{2i}x_{2j})^2$

= $[x_{1i}x_{1j} + x_{2i}x_{2j}]^2$
= $[(x_{1i}, x_{2i}) \bullet (x_{1j}, x_{2j})]^2$
= $(\mathbf{x_i} \bullet \mathbf{x_j})$

So to compare 2 of the original vectors, we just take their dot product and square it. In general, we develop functions of dot products that are useful in classification.