1) Sigmoid derivative

$$\sigma'(z) = \frac{d\sigma(z)}{dz} = \frac{d}{dz}\frac{1}{1+e^{-z}} = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1+e^{-z}-1}{(1+e^{-z})^2} =$$
$$= \frac{1}{1+e^{-z}} - \frac{1^2}{(1+e^{-z})^2} = \sigma(z) - \sigma^2(z) = \sigma(z)(1-\sigma(z))$$

***2) Prove, that** $\sigma(-z) = 1 - \sigma(z)$*

$$\sigma(-z) + \sigma(z) = \frac{1}{1+e^{z}} + \frac{1}{1+e^{-z}} = \frac{2+e^{z}+e^{-z}}{2+e^{z}+e^{-z}} = 1$$

3) Calculate the loss function derivative:

$$L(w, x_1, \ldots, x_N) = -\frac{1}{N}\sum_{i=1}^{N}[y_i\ log(\sigma(w^T x_i)) + (1-y_i)\ log(1-\sigma(w^T x_i))] + \alpha\sum_{j=1}^{M} w_j^2$$

$$\nabla_w L(w, X) = \frac{1}{N}\sum_{i=1}^{N} y_i \frac{1}{\sigma(w^T x_i)}\sigma(w^T x_i)(1-\sigma(w^T x_i))\ x_i -$$

$$-(1-y_i)\frac{1}{1-\sigma(w^T x_i)}\sigma(w^T x_i)(1-\sigma(w^T x_i))\ x_i + 2\alpha w =$$
$$= \frac{1}{N}\sum_{i=1}^{N} x_i\ (\sigma(w^T x_i) - y_i) + 2\alpha w,$$

where x_i - is i-th object from training set

4) Weights update rule

Take a step in direction, opposite to the direction of the gradient $\nabla_w L$:

$$w_{n+1} = w_n - \gamma\nabla_w L$$

5) Proof of loss convexity

Sufficient conditions for convexity of a function:
If $f''(x) \geq 0$ for any x from $[a, b]$, then $f(x)$ - is convex function on $[a, b]$.
So, the calculation of the second derivative of the loss function gives us formula:

$$\nabla_w \nabla_w L = 2\alpha + \frac{1}{N} \sum_{i=1}^{N} x_i \, x_i(\sigma - \sigma^2) \geq 0,$$

where x_i - is i-th object from training set. The second derivation w.r.t. w is always great or equal than zero thus $\sigma - \sigma^2 \geq 0$ bacause $\sigma(x) \leq 1$ for any x.

In [8]:
```
from matplotlib import pyplot as plt
from numpy import linspace
X = linspace(0, 1, 200)
plt.plot(X, (lambda x: x)(X), X, (lambda x: x**2)(X))
plt.show()
```