Let $(\mathbf{x}_i, y_i)$ denote the training data pair of feature vectors and labels. The regularized logistic regression max log-likelihood can be written as

$$\min_{\mathbf{w},b} E = \frac{\lambda}{2}\|\mathbf{w}\|^2 + \sum_i g(-y_i(\mathbf{w}^T\mathbf{x}_i + b)),$$
$$g(\psi) = \log(1 + e^{\psi}).$$

Note that $g(\psi)$ is the negative of the log-likelihood associated with the probability model,

$$p(y|\mathbf{x}) = \frac{1}{1 + \exp(-y\mathbf{w}^T\mathbf{x} + b)}.$$

We can rewrite $E$ as the constrained minimization

$$\min E = \frac{\lambda}{2}\|\mathbf{w}\|^2 + \sum_i g(\psi_i),$$
$$\psi_i = -y_i(\mathbf{w}^T\mathbf{x}_i + b),\ \forall i.$$

Thus, the Lagrangian for this problem is

$$L = \frac{\lambda}{2}\|\mathbf{w}\|^2 + \sum_i g(\psi_i) + \sum_i \alpha_i \left[-\psi_i - y_i(\mathbf{w}^T\mathbf{x}_i + b)\right],$$

which has the KKT optimality conditions given by

$$\nabla_{\mathbf{w}} L = \lambda\mathbf{w} - \sum_i \alpha_i y_i \mathbf{x}_i = 0;$$
$$\frac{\partial L}{\partial b} = \sum_i \alpha_i y_i = 0;$$
$$\frac{\partial L}{\partial \psi_i} = g'(\psi_i) - \alpha_i = 0,\ \forall i.$$

Hence, these conditions give

$$\mathbf{w}(\alpha) = \frac{1}{\lambda}\sum_i \alpha_i y_i \mathbf{x}_i; \quad \psi_i(\alpha_i) = (g')^{-1}(\alpha_i),$$

and the constraint $\sum_i \alpha_i y_i = 0$.

To continue, let $G(\alpha_i) = g(\psi_i) - \alpha_i\psi_i$, where this is part of $L$, then

$$\frac{\partial G}{\partial \alpha_i} = \frac{\partial \psi_i}{\partial \alpha_i} g'(\psi_i) - \alpha_i \frac{\partial \psi_i}{\partial \alpha_i} - \psi_i = -\psi_i = -(g')^{-1}(\alpha_i).$$

Therefore, $G'(\alpha_i) = -(g')^{-1}(\alpha_i)$. We can now substitute in the logistic regression probability model and find $(g')^{-1}(z)$,

$$\begin{aligned}
g(z) &= \log(1+e^z), \\
g'(z) &= e^z/(1+e^z), \\
g'(z)(1+e^z) &= e^z, \\
g'(z) &= e^z(1-g'(z)) \\
g'(z)/(1-g'(z)) &= e^z \\
\log(g'(z)/(1-g'(z))) &= z = (g')^{-1}(g'(z)).
\end{aligned}$$

Thus,

$$\begin{aligned}
(g')^{-1}(u) &= \log(u/(1-u)), \\
G(\alpha_i) &= -\alpha_i \log \alpha_i - (1-\alpha_i)\log(1-\alpha_i).
\end{aligned}$$

This gives us the new objective

$$\begin{aligned}
\min_{\alpha,b} L &= \frac{\lambda}{2}\|\mathbf{w}(\alpha)\|^2 + \sum_i G(\alpha_i) - \sum_i \alpha_i y_i(\mathbf{w}(\alpha)^T\mathbf{x}_i + b), \\
\text{s.t. } & \sum_i \alpha_i y_i = 0.
\end{aligned}$$

Due to the constraint, the term $\sum_i \alpha_i y_i b = 0$. Note that this is also why you cannot just append a column of ones to the original feature vector to get the bias implicitly—this weight will just go to zero. Then, we can substitute in $\mathbf{w}(\alpha)$ and get

$$L = \frac{1}{2\lambda}\sum_{i,j}\alpha_i\alpha_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j + \sum_i G(\alpha_i) - \frac{1}{\lambda}\sum_{i,j}\alpha_i\alpha_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j.$$

Combining terms and putting it all together with $K = [\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)]^{n\times n}$, we get

$$\min_\alpha E(\alpha) = -\frac{1}{2\lambda}(\alpha\circ\mathbf{y})^T K(\alpha\circ\mathbf{y}) + \sum_i G(\alpha_i),$$

where $\circ$ is the Hadamard product (.* in Matlab). This provides the optimal $\alpha$, which is used to compute $\mathbf{w}(\alpha)$, but does not compute $b$, which is essential. However, remember that we are maximizing the log-likelihood of the logistic probability model,

$$p(y|\mathbf{x}) = \frac{1}{1+\exp(-y(\mathbf{w}(\alpha)^T\mathbf{x}+b)}.$$

Hence, we can substitute in $\mathbf{w}(\alpha)^T\mathbf{x}_j = \sum_i \alpha_i y_i \mathbf{x}_i^T\mathbf{x}_j = (\alpha\circ\mathbf{y})^T K_j$, where $K_j$ is the $j$th column of $K$. Then, we can maximize the log-likelihood wrt $b$,

$$\max_b \log\left(\prod_j p(y_j|\mathbf{x}_j)\right) = \min_b \sum_j \log\left(1+\exp(-y_j((\alpha\circ\mathbf{y})^T K_j + b)\right).$$

Note that optimization objective can be vectorized in Matlab as

$$\operatorname{sum}\log(1+\exp(-\mathbf{y}.*(K(\alpha\circ\mathbf{y})+b))).$$

This is an unconstrained optimization; hence, you can use `fminunc`.