Spring 2013

(3.1) We are faced with a content filtering problem where the idea is to rank new songs by trying to predict how they might be rated by a particular user. Each song x is represented by a feature vector $\phi(x)$ whose coordinates capture specific acoustical properties. The ratings are binary valued $y \in \{0, 1\}$ ("need earplugs" or "more like this"). Given n already rated songs, we decided to use regularized linear regression to predict the binary ratings. The training criterion is

$$J(\theta) = \frac{\lambda}{2}\|\theta\|^2 + \frac{1}{n}\sum_{t=1}^{n}(y^{(t)} - \theta \cdot \phi(x^{(t)}))^2/2 \tag{3}$$

(a) **(8 points)** Let $\hat{\theta}$ be the optimal setting of the parameters with respect to the above criterion, which of the following conditions must be true (check all that apply)

( ) $\lambda\,\hat{\theta} - \frac{1}{n}\sum_{t=1}^{n}(y^{(t)} - \hat{\theta} \cdot \phi(x^{(t)}))\phi(x^{(t)}) = 0$

( ) $J(\hat{\theta}) \geq J(\theta)$, for all $\theta \in \mathcal{R}^d$

( ) If we increase λ, the resulting $\|\hat{\theta}\|$ will decrease

( ) If we add features to $\phi(x)$ (whatever they may be), the resulting squared training error will NOT increase

(b) **(3 points)** Once we have the estimated parameters $\hat{\theta}$, we must decide how to predict ratings for new songs. Note that the possible rating values are 0 or 1. When do we choose rating $y = 1$ for a new song x? Please write the corresponding expression.

(c) **(3 points)** If we change λ, we obtain different $\hat{\theta}$, and therefore different rating predictions according to your rule above. What will happen to your predicted ratings when we increase the regularization parameter λ?