

Problem 1)

$$\begin{aligned} X_n &\in \mathbb{R}^d & X_n &\sim N(Wz_n, \sigma^2 I) \\ W &\in \mathbb{R}^{d \times K} & W &\sim N(0, \lambda^{-1} I) \\ Z_n &\in \mathbb{R}^K & Z_n &\sim N(0, I) \end{aligned}$$

$$p(X | Wz_n, \sigma^2 I) = (2\pi\sigma^2 I)^{-\frac{d}{2}} \exp\left[-\frac{1}{2\sigma^2} (X_n - Wz_n)^T (X_n - Wz_n)\right]$$

$$p(W) = \left(\frac{\lambda}{2\pi}\right)^{\frac{dK}{2}} \exp\left[-\frac{\lambda}{2} \text{trace}(W^T W)\right]$$

$$p(z_n) = (2\pi)^{-\frac{K}{2}} \exp\left[-\frac{1}{2} z_n^T z_n\right]$$

In order to find

$$W' = \arg \max_W \ln p(x_1, \dots, x_n, W)$$

Note that

$$\ln p(x_1, \dots, x_n, W) = \int q(z_n) \ln \frac{p(x_1, \dots, x_n, W, z_n)}{q(z_n)} dz_n$$

$$q(z) = p(z | W, X) \propto \prod_{i=1}^n p(x_i | W, z_i) p(z_i) p(W | z_i)$$

$$\begin{aligned} q(z) &\propto \prod_{i=1}^n (2\pi\sigma^2 I)^{-\frac{d}{2}} \exp\left[-\frac{1}{2\sigma^2} (X_i - Wz_i)^T (X_i - Wz_i)\right] (2\pi)^{-\frac{K}{2}} \exp\left[-\frac{1}{2} z_i^T z_i\right] \\ &= (2\pi\sigma^2 I)^{-\frac{dn}{2}} (2\pi)^{-\frac{nK}{2}} \exp\left[-\frac{1}{2\sigma^2} (X_n - Wz_n)^T (X_n - Wz_n) - \frac{1}{2} z_n^T z_n\right] \\ &= (2\pi)^{-\frac{n(d+K)}{2}} (\sigma^2 I)^{-\frac{dn}{2}} \exp\left[\sum_{i=1}^n -\frac{1}{2\sigma^2} [X_i^T X_i - X_i^T Wz_i - (Wz_i)^T X_i + (Wz_i)^T Wz_i] - \frac{1}{2} z_i^T z_i\right] \\ &= \text{constants} \exp\left[-\frac{1}{2} \sum_{i=1}^n \left[\frac{X_i^T X_i}{\sigma^2} - \frac{X_i^T Wz_i}{\sigma^2} - \frac{X_i^T Wz_i}{\sigma^2} + \frac{(Wz_i)^T Wz_i}{\sigma^2} + \frac{z_i^T z_i}{\sigma^2} \right]\right] \\ &\propto \exp\left[-\frac{1}{2} \sum_{i=1}^n \left[\frac{X_i^T X_i}{\sigma^2} - \frac{2X_i^T Wz_i}{\sigma^2} + \frac{(Wz_i)^T Wz_i}{\sigma^2} + \frac{z_i^T z_i}{\sigma^2} \right]\right] \\ &= \exp\left[-\frac{1}{2} \sum_{i=1}^n \left[\frac{X_i^T X_i}{\sigma^2} - \frac{2X_i^T Wz_i}{\sigma^2} + \frac{z_i^T W^T W z_i}{\sigma^2} + \frac{z_i^T z_i}{\sigma^2} \right]\right] \\ &= \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i^T X_i - 2X_i^T Wz_i + z_i^T W^T W z_i + z_i^T z_i)\right] \\ &= \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i^T X_i - 2X_i^T Wz_i + z_i^T W^T W z_i + z_i^T z_i)\right] \end{aligned}$$

$$q(z) \propto \exp \left[-\frac{\text{tr}(W^T W) + \sigma^2}{2\sigma^2} \sum_{n=1}^N \left(\frac{X_n^T X_n}{\text{tr}(W^T W) + \sigma^2} - \frac{2X_n^T W z_n}{\text{tr}(W^T W) + \sigma^2} + z_n^T z_n \right) \right]$$

depend on z

$$\propto \exp \left[-\frac{\text{tr}(W^T W) + \sigma^2}{2\sigma^2} \sum_{n=1}^N \left(z_n^T - \frac{X_n^T W}{\text{tr}(W^T W) + \sigma^2} \right) \left(z_n - \frac{X_n^T W}{\text{tr}(W^T W) + \sigma^2} \right) \right]$$

which is a Normal distribution with

$$p(z|w, x) \sim N \left(\frac{\sum_{n=1}^N X_n^T W}{\text{tr}(W^T W) + \sigma^2 I}, \frac{\sigma^2 I}{N(\text{tr}(W^T W) + \sigma^2 I)} \right)$$

Then, we need to take the expectation of the joint likelihood and maximize it.

$$\mathcal{L}(w) = E_q[\ln p(z, x, w)]$$

$$p(z, x, w) \propto p(w) \prod_{n=1}^N p(x_n | w, z_n) p(z_n)$$

$$\ln p(z, x, w) \propto \ln \left[\left(\frac{\lambda}{2\pi} \right)^{\frac{d}{2}} \exp \left[-\frac{\lambda}{2} \text{trace}(W^T W) \right] (2\pi\sigma^2)^{-\frac{N}{2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=1}^N (X_n - W z_n)^T (X_n - W z_n) \right] \right]$$

$$\times (2\pi)^{-\frac{N}{2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=1}^N z_n^T z_n \right]$$

$$= \ln \left[\left(\frac{\lambda}{2\pi} \right)^{\frac{d}{2}} (2\pi)^{-\frac{N}{2}} (\sigma^2)^{-d} \exp \left[-\frac{\lambda}{2} \text{trace}(W^T W) - \frac{1}{2\sigma^2} \sum_{n=1}^N (X_n - W z_n)^T (X_n - W z_n) - \frac{1}{2\sigma^2} \sum_{n=1}^N z_n^T z_n \right] \right]$$

$$= \ln \left[\lambda^{\frac{d}{2}} (2\pi)^{-\frac{N}{2}} \sigma^{-d} \exp \left[-\frac{\lambda}{2} \text{trace}(W^T W) - \frac{1}{2\sigma^2} \sum_{n=1}^N (X_n - W z_n)^T (X_n - W z_n) - \frac{1}{2\sigma^2} \sum_{n=1}^N z_n^T z_n \right] \right]$$

$$= \underbrace{\left(\frac{d}{2} \ln \lambda - \frac{N}{2} \ln 2\pi - d \ln \sigma \right)}_{\text{constants wrt } w} - \frac{\lambda}{2} \text{trace}(W^T W) - \frac{1}{2\sigma^2} \sum_{n=1}^N (X_n - W z_n)^T (X_n - W z_n) - \frac{1}{2\sigma^2} \sum_{n=1}^N z_n^T z_n$$

$$= \text{constants} - \frac{\lambda}{2} \text{trace}(W^T W) - \frac{1}{2\sigma^2} \sum_{n=1}^N X_n^T X_n + \frac{1}{\sigma^2} \sum_{n=1}^N X_n^T W z_n - \frac{1}{2\sigma^2} \text{trace}(W^T W) \sum_{n=1}^N z_n^T z_n$$

$$\mathcal{L}(w) = \text{constants} - \frac{\lambda}{2} \text{trace}(W^T W) - \frac{1}{2\sigma^2} \sum_{n=1}^N X_n^T X_n + \frac{1}{\sigma^2} \sum_{n=1}^N X_n^T W E[z_n] - \frac{1}{2\sigma^2} \text{trace}(W^T W) \sum_{n=1}^N E[z_n^T z_n]$$

$$\nabla_w \mathcal{L}(w) = 0 = -\lambda W - \frac{W}{\sigma^2} \sum_{n=1}^N E[z_n z_n^T] + \sum_{n=1}^N X_n E[z_n]$$

$$W \left(\lambda + \frac{1}{\sigma^2} \sum_{n=1}^N E[z_n z_n^T] \right) = \sum_{n=1}^N X_n E[z_n]$$

$$W_{\text{MSE}}^* = \frac{\sum_{n=1}^N X_n E[z_n]}{\lambda + \frac{1}{\sigma^2} \sum_{n=1}^N E[z_n z_n^T]} = \frac{\sum_{n=1}^N X_n \mu^*}{\lambda + \frac{1}{\sigma^2} \Sigma^*}$$

Thus, we have both the E-step & the M-step. The next step that is left is to verify convergence by calculating $\ln p_t(x_1, \dots, x_n, w)$ and ensuring that as $t \rightarrow \infty$ the term is monotonically increasing. To calculate $\ln p_t$,

$$\ln p_t(x_1, \dots, x_n, w) = \ln \int q(z_n) \frac{p(x_1, \dots, x_n, w, z_n)}{q(z_n)} dz_n.$$

$$= E_z [\ln p(x_1, \dots, x_n, w, z_n)]$$

$$= -\frac{\lambda}{2} \text{trace}(W^T W) - \frac{1}{2\sigma^2} \sum_{n=1}^N x_n^T x_n + \frac{1}{2\sigma^2} \sum_{n=1}^N x_n^T W \mu^* - \frac{1}{2\sigma^2} \text{trace}(W W^T) \sum_{n=1}^N \sum^*$$

$$\ln p_{t+1}(x_1, \dots, x_n, w_t) = -\frac{\lambda}{2} \text{trace}(W_t^T W_t) - \frac{1}{2\sigma^2} \sum_{n=1}^N x_n^T x_n + \frac{1}{2\sigma^2} \sum_{n=1}^N x_n^T W_t \left(\frac{\sum_{n=1}^N x_n^T W_t}{\text{trace}(W_t^T W_t) + \sigma^2 I} \right) - \frac{1}{2\sigma^2} \text{trace}(W_t W_t^T) N \left(\frac{\sum_{n=1}^N x_n^T W_t}{N \cdot \text{trace}(W_t^T W_t) + \sigma^2 I} \right)$$

This EM algorithm can be implemented by the following:

1. Initialize w_0 as a $(d \times k)$ matrix with zeros.

2. For iterations $t=1, 2, \dots, T$

a. Calculate $E[z] = \frac{N^{-1} \sum x_i^T W}{\text{trace}(W^T W) + \sigma^2 I}$ (E-step)

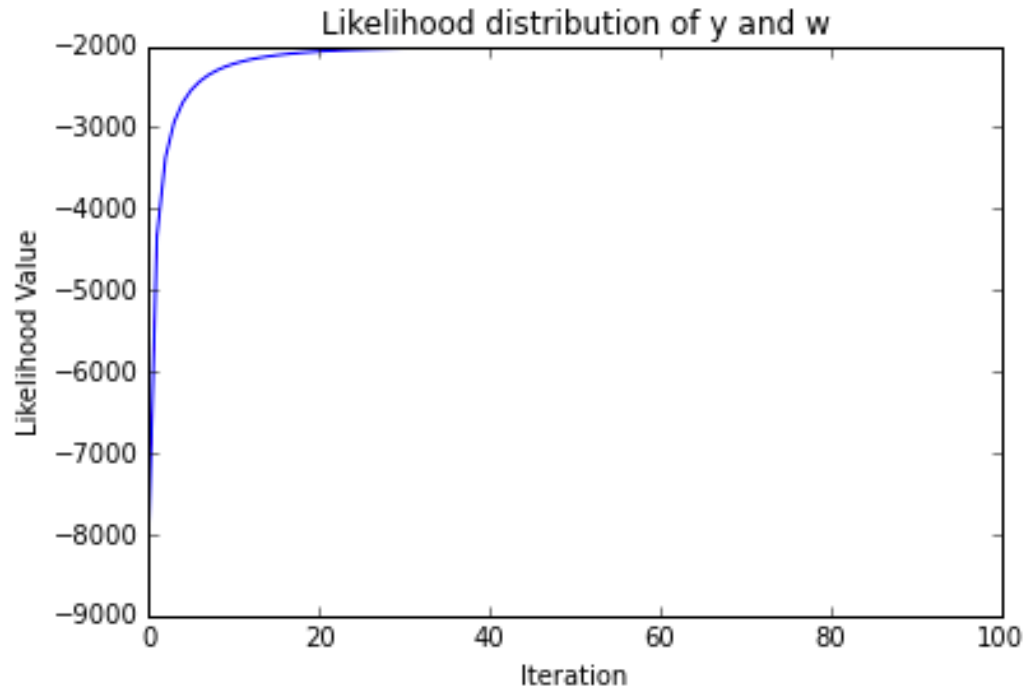
b. Update w_t^* as $w_{t+1} = \frac{\sum_{n=1}^N x_n E(z_n)}{\lambda + \sigma^2 E(z_n z_n^T)}$ (M-step).

c. Calculate $\ln p_t(x_1, \dots, x_n, w)$ and stop when $\ln p_t(\cdot) - \ln p_{t-1}(\cdot) < \epsilon$ for some small $\epsilon > 0$.

Problem 2

a. See Python code attached.

b.

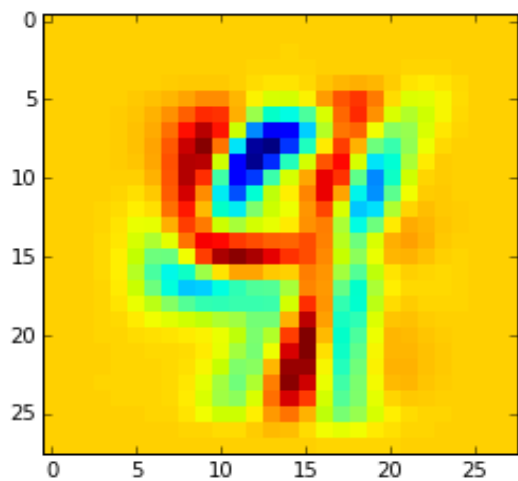


c.

	Actual = 4	Predicted = 9
Predicted = 4	930	52
Predicted = 9	77	932

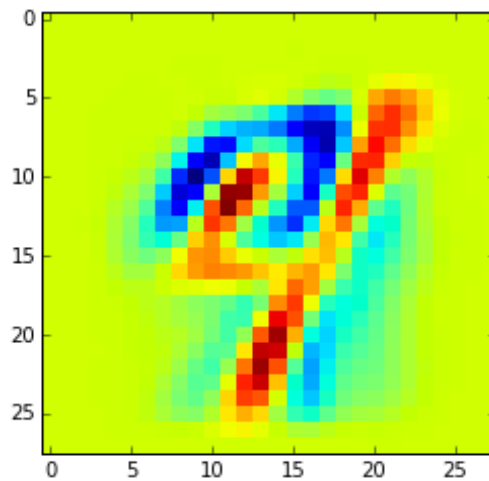
d. Misclassified digits

Observation 40



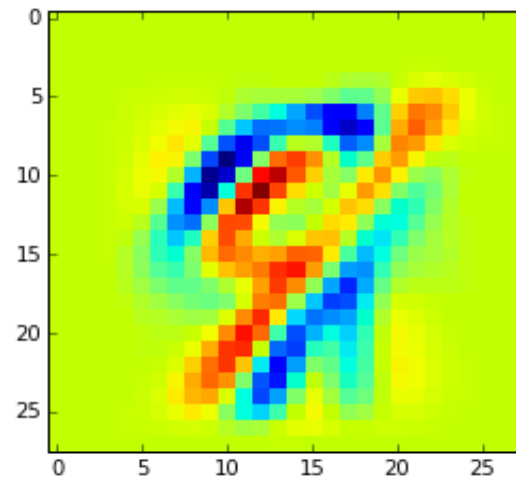
Predicted Class = 9
Actual Class = 4
 $E[Y=9|X] = 0.754821$

Observation 46



Predicted Class = 9
Actual Class = 4
 $E[Y=9|X] = 0.78199$

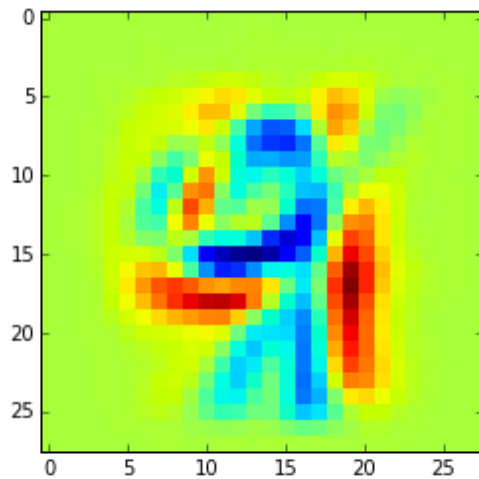
Observation 340



Predicted Class = 9
Actual Class = 4
 $E[Y=9|X] = 0.50465$

e. Most ambiguous predictions

Observation 586

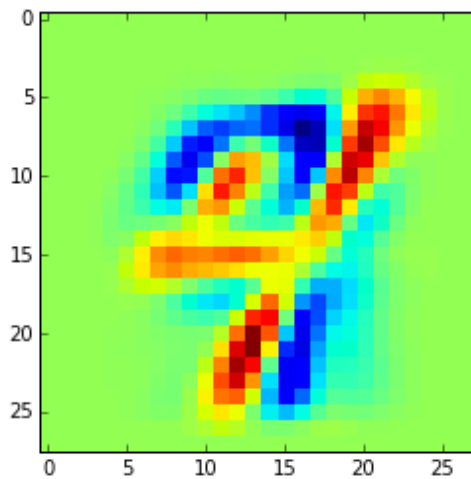


Predicted Class = 9

Actual Class = 4

$$E[Y=9|X] = 0.50026$$

Observation 340

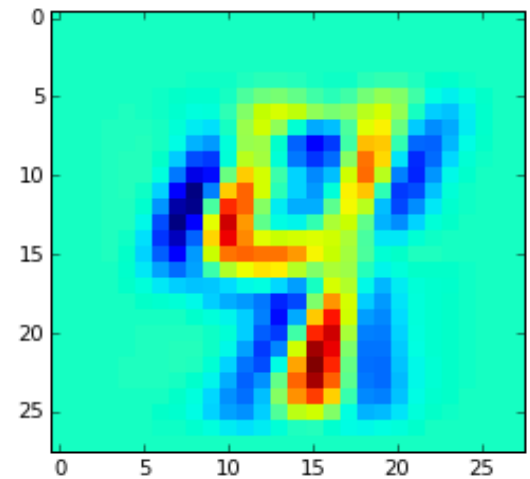


Predicted Class = 9

Actual Class = 4

$$E[Y=9|X] = 0.50465$$

Observation 210

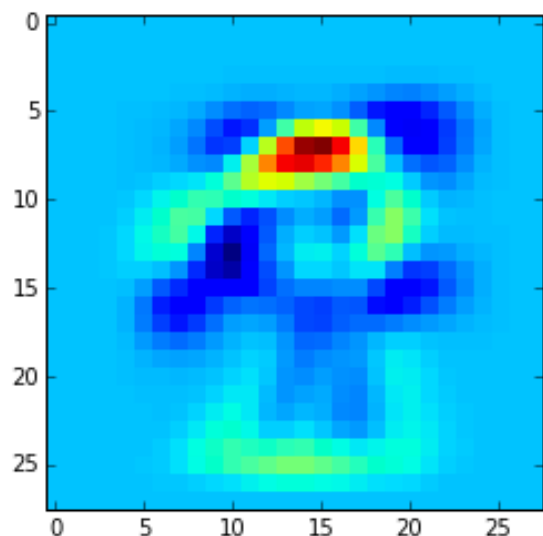


Predicted Class = 9

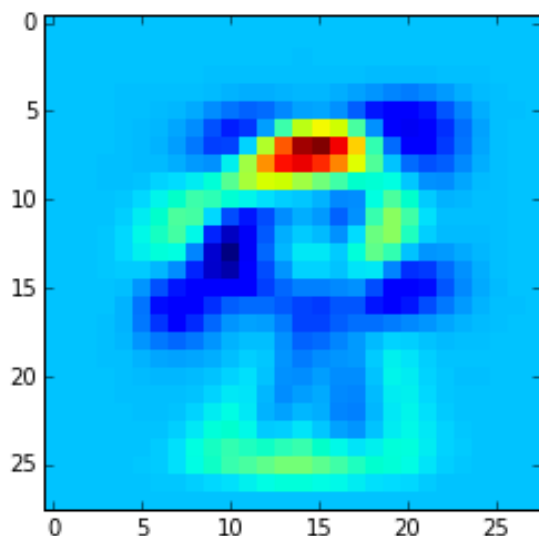
Actual Class = 4

$$E[Y=9|X] = 0.506176$$

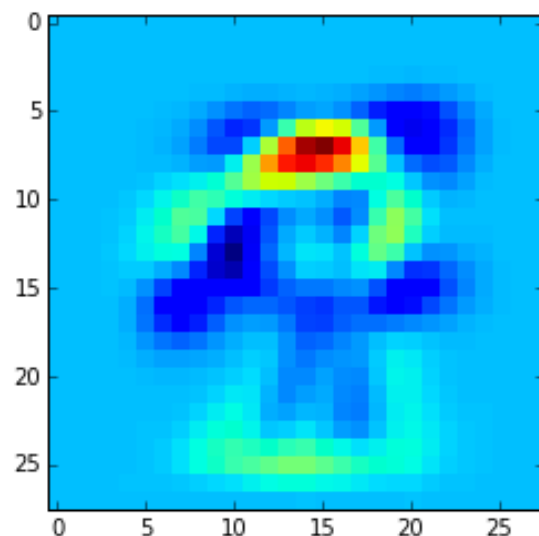
f. Reconstructed w_t



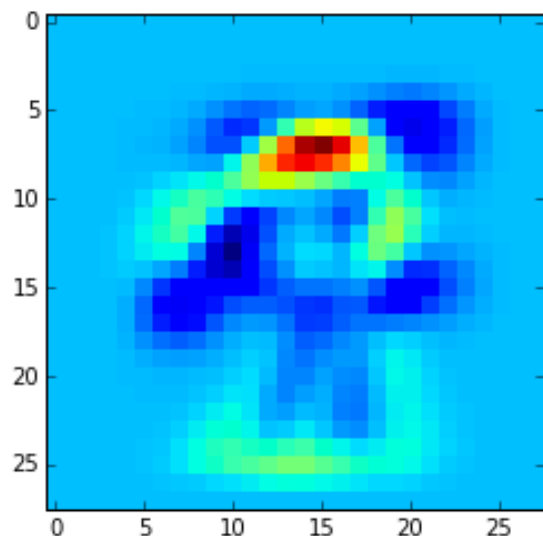
$t = 1$



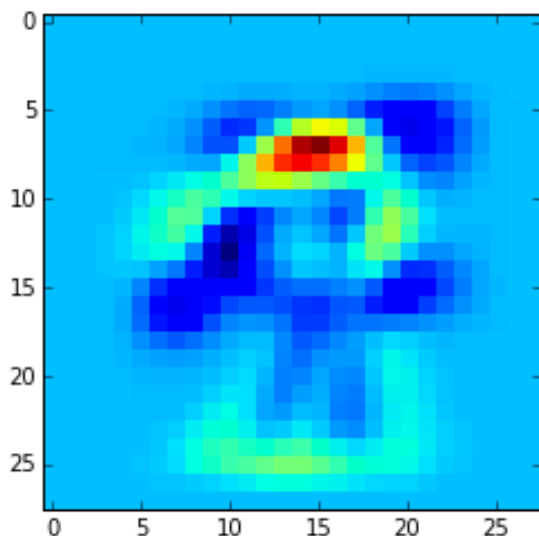
$t = 5$



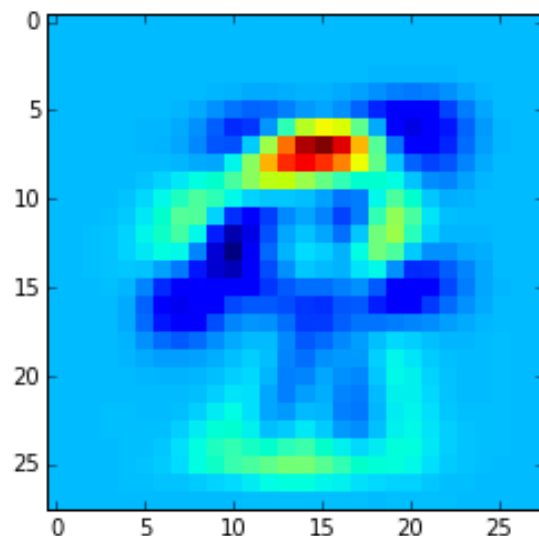
$t = 10$



$t = 25$



$t = 50$



$t = 100$