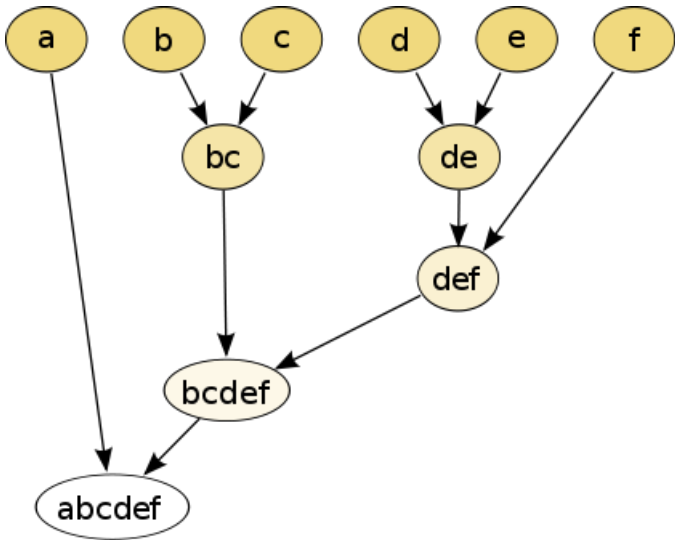
Unsupervised learning

Hierarchical clustering

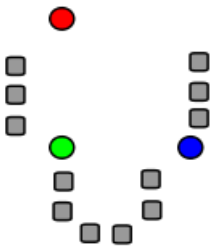


Example

Run single-link distance based hierarchical clustering:

	1	2	3	4	5	6
1	0	4	13	24	12	8
2		0	10	22	11	10
3			0	7	3	9
4				0	6	18
5					0	8.5
6						0

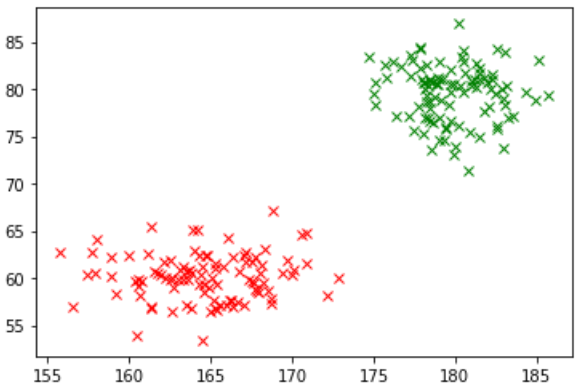
Sum-of-squares Methods



In [18]:
```
import numpy as np
import matplotlib.pyplot as plt

# Some random experiments with 2D Gaussians
mu1 = [165,60]
cov1 = [[10,0],[0,5]]
mu2 = [180,80]
cov2 = [[6,0],[0,10]]
x1 = np.random.multivariate_normal(mu1, cov1, 100)
x2 = np.random.multivariate_normal(mu2, cov2, 100)
plt.plot(x1[:,0],x1[:,1],'rx')
plt.plot(x2[:,0],x2[:,1],'gx')
```

Out[18]:
```
[<matplotlib.lines.Line2D at 0x7f5f7e121be0>]
```

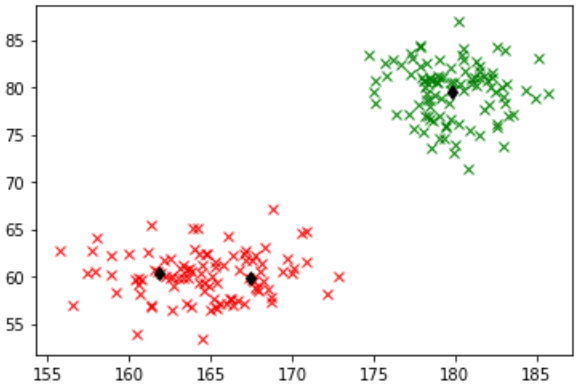


In [19]:
```
# Let's test k-means clustering
from sklearn.cluster import KMeans

kmeans = KMeans(init="random",n_clusters=3,n_init=1,max_iter=10)
X = np.row_stack((x1, x2))
kmeans.fit(X)
print(kmeans.cluster_centers_)
plt.plot(x1[:,0],x1[:,1],'rx')
plt.plot(x2[:,0],x2[:,1],'gx')
plt.plot(kmeans.cluster_centers_[:,0],kmeans.cluster_centers_[:,1],'kd')
```

```
[[179.78244427  79.56641538]
 [161.85265932  60.38321698]
 [167.48526126  59.88957398]]
```

Out[19]:
```
[<matplotlib.lines.Line2D at 0x7f5f7e1016a0>]
```



In [20]:
```
# K-means clustering of Cifar-10 images (10 clusters) to see how well unsuperv
ised learning
# works against supervised
```

References

A.R. Webb and K.D. Copsey (2011), Statistical Pattern Recognition, 3rd Edition, Chapter 11.