Claims

1. A method of automatically assigning class labels to objects, using object data indicative of a plurality of parameters associated with each object, the method comprising:
   (i) identifying, from the object data or from a lower-dimensional encoding of the object data, a plurality of cluster centres in a d-dimensional space, each cluster centre corresponding to one of the class labels;
   (ii) for respective cluster centres, determining a surrounding region based on a nearest neighbor cluster centre, and assigning the respective class label to objects within the surrounding region;
   (iii) generating a predictive model using the object data, or the lower-dimensional encoding of the object data, and the class labels of the assigned objects; and
   (iv) assigning the class labels of the assigned objects to unassigned objects using the predictive model.

2. A method according to claim 1, wherein the cluster centres are identified by: determining a kernel density estimate from the object data; and detecting peaks in the kernel density estimate, said peaks corresponding to the cluster centres.

3. A method according to claim 1 or claim 2, further comprising, prior to operation (i), applying dimensionality reduction to the object data to generate the lower-dimensional encoding of the object data.

4. A method according to claim 3, wherein after the dimensionality reduction, the lower-dimensional encoding of the object data defines a 2-dimensional space.

5. A method according to any one of the preceding claims, wherein the surrounding region is determined by determining a distance dk to the nearest neighbor cluster centre, and wherein the surrounding region is a d-ball of radius less than or equal to $dk/2$ centred on the cluster centre.

6. A method according to any one of claims 2 to 5, comprising optimizing the kernel bandwidth *H* for the kernel density estimation.

7. A method according to claim 6, wherein *H* is optimized by minimizing the asymptotic mean integrated standard error (AMISE) of the kernel density estimate.

8. A method according to any one of the preceding claims, wherein the object data is flow cytometry data or mass cytometry data, and wherein the objects are cells.

9. A method according to claim 8, wherein the plurality of parameters comprises expression levels for a plurality of proteins.

10. A computer system for automatically assigning class labels to objects, using object data indicative of a plurality of parameters associated with each object, the computer system comprising at least one processor and a data storage device storing program instructions, the program instructions being operative, upon being run by the processor to cause the processor to perform which is configured to:
    (i) identify, from the object data or from a lower-dimensional encoding of the object data, a plurality of cluster centres in a d-dimensional space, each cluster centre corresponding to one of the class labels;
    (ii) for respective cluster centres, determine a surrounding region based on a nearest neighbor cluster centre, and assigning the respective class label to objects within the surrounding region;
    (iii) generate a predictive model using the object data, or the lower-dimensional encoding of the object data, and the class labels of the assigned objects; and
    (iv) assign the class labels of the assigned objects to unassigned objects using the predictive model.

11. A computer system according to claim 10, wherein the data storage device stores program instructions operative upon being run by the processor to cause the processor

to identify the cluster centres by: determining a kernel density estimate from the object data; and detecting peaks in the kernel density estimate, said peaks corresponding to the cluster centres.

12. A computer system according to claim 10 or claim 11, wherein the data storage device stores program instructions operative upon being run by the processor to cause the processor to, prior to operation (i), apply dimensionality reduction to the object data to generate the lower-dimensional encoding of the object data.

13. A computer system according to claim 12, wherein after the dimensionality reduction, the lower-dimensional encoding of the object data defines a 2-dimensional space.

14. A computer system according to any one of the claims 10 to 13, wherein the data storage device stores program instructions operative upon being run by the processor to cause the processor to determine the surrounding region by determining a distance *dk* to the nearest neighbor cluster centre, and wherein the surrounding region is a d-ball of radius less than or equal to *dk*/2 centred on the cluster centre.

15. A computer system according to any one of claims 11 to 14, wherein the data storage device stores program instructions operative upon being run by the processor to cause the processor to optimize the kernel bandwidth *H* for the kernel density estimation.

16. A computer system according to claim 15, wherein the data storage device stores program instructions operative upon being run by the processor to cause the processor to optimize *H* by minimizing the asymptotic mean integrated standard error (AMISE) of the kernel density estimate.

17. A computer system according to any one of claims 10 to 16, wherein the object data is flow cytometry data or mass cytometry data, and wherein the objects are cells.

18. A computer system according to claim 17, wherein the plurality of parameters comprises expression levels for a plurality of proteins.

19. A non-transitory computer-readable medium having stored thereon computer program instructions which are configured to, when executed by at least one processor, perform the method of any one of claims 1 to 9.