Supplementary material: Supervised nonparametric classification in the context of replicated point patterns

Image Analysis & Stereology

Kateřina Pawlasová, Jiří Dvořák
Department of Probability and Mathematical Statistics,
Faculty of Mathematics and Physics, Charles University,
Sokolovská 83, 186 75 Praha 8, Czech Republic
dvorak@karlin.mff.cuni.cz

S1 Nonparametric estimation of pair correlation function

This section briefly looks at the nonparametric estimation of the pair correlation function. We follow the book [6, Sect. 4.3], where a comprehensive list of functional summary characteristics and their estimators can be found.

Suppose that a stationary point process X is observed in a compact observation window W. First, a natural unbiased estimator of the intensity λ is given by $\widehat{\lambda} = N_X(W)/|W|$. For the Poisson point process, it is the maximum likelihood estimator.

Suppose now that X is stationary and isotropic, which implies that g is a function of a scalar argument $r > 0$. Let us take a probability density function s on $\mathbb{R}$ and a constant $\beta > 0$. For $x \in \mathbb{R}$, set $s_\beta(x) = (1/\beta)s(x/\beta)$. The nonparametric kernel estimator of g with edge correction factor e_W^r is then given by

$$\widehat{g}(r) = \sum_{x,y \in X \cap W}^{\neq} \frac{s_\beta\left(r - \|x - y\|\right)}{\widehat{\lambda^2}\pi r|W|} e_W^r(x, y),$$

where $\widehat{\lambda^2} = \left(N_X(W)\left(N_X(W) - 1\right)\right)/|W|^2$. Various choices of e_W^r are available [5, Sect. 4.2.2]. This kernel estimator is implemented in the function `pcf.ppp` of the `R` package `spatstat` [2, 7].

S2 Experiment 1

This section provides additional visual material accompanying Experiment 1 form the main text. Binary classification in the sense of "clustering/repulsive interactions vs complete spatial randomness" is considered.

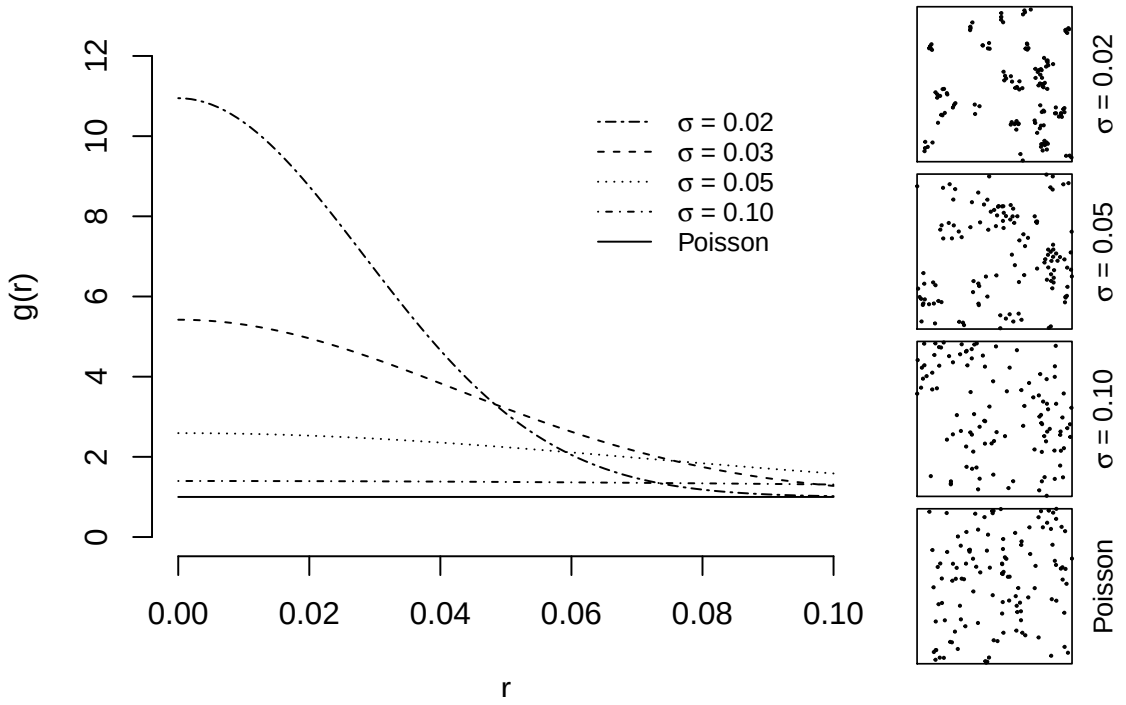

Figure S2.1: Illustration of the dependence of the behaviour of $\Phi(\sigma)$ on the model parameter σ. For selected values of σ, a realization of $\Phi(\sigma)$ is plotted as well as the evolution of the values of the pair correlation function g. The Poisson point process Π can be viewed as the limit of $\Phi(\sigma)$ for $\sigma \to \infty$.

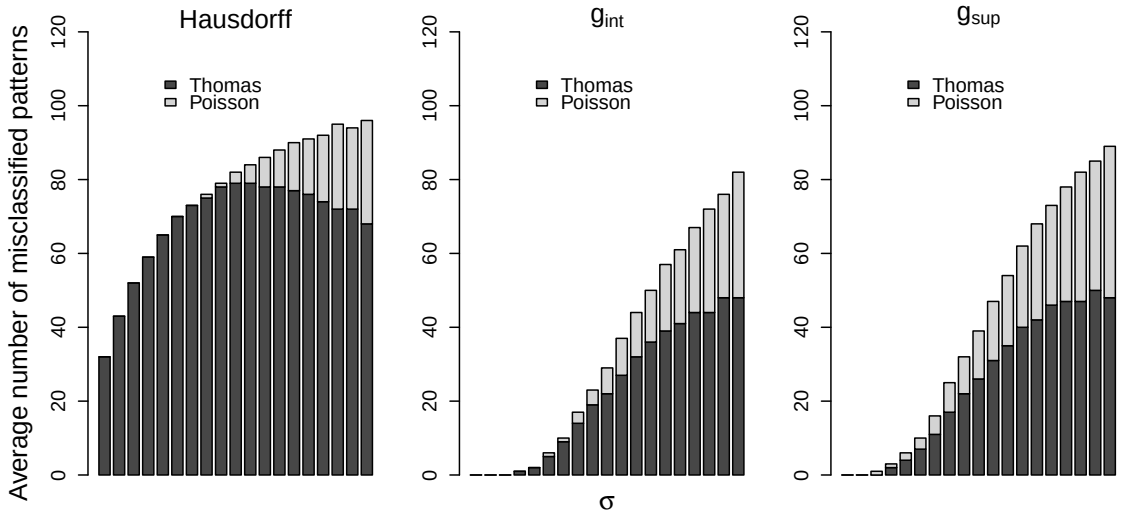

Figure S2.2: Binary classification $\Phi(\sigma)$ vs Π. For each $\sigma \in \Sigma = \{0.02, 0.03, \ldots, 0.20\}$, the average number (from the 100 repetitions) of misclassified patterns in the testing data is plotted. We distinguish between realizations of the Thomas process $\Phi(\sigma)$ (dark grey) and the Poisson point process Π (light grey). The plots correspond to d_H (left), $\mathrm{d}_{int}[g]$ (middle) and $\mathrm{d}_{sup}[g]$ (right). For the Hausdorff measure, realizations of $\Phi(\sigma)$ form the majority of the misclassified patterns, disregarding the value of σ. Note that $\Phi(\sigma)$ shows a higher variability in terms of realizations than Π, especially for smaller values of σ; for further comments see Experiment 1 in the main text. For $\mathrm{d}_{int}[g]$, only a small amount of realizations of $\Phi(\sigma)$ is misclassified for small values of σ. As σ grows, the number of misclassified realizations of $\Phi(\sigma)$ is growing and misclassified realizations of Π start to appear. For the largest values of σ (weak clustering interactions, realizations of $\Phi(\sigma)$ barely distinguishable from realizations of Π on the unit square), we observe almost the same number of misclassified realizations for both of the two groups. Similar observations hold for the dissimilarity measure $\mathrm{d}_{sup}[g]$.

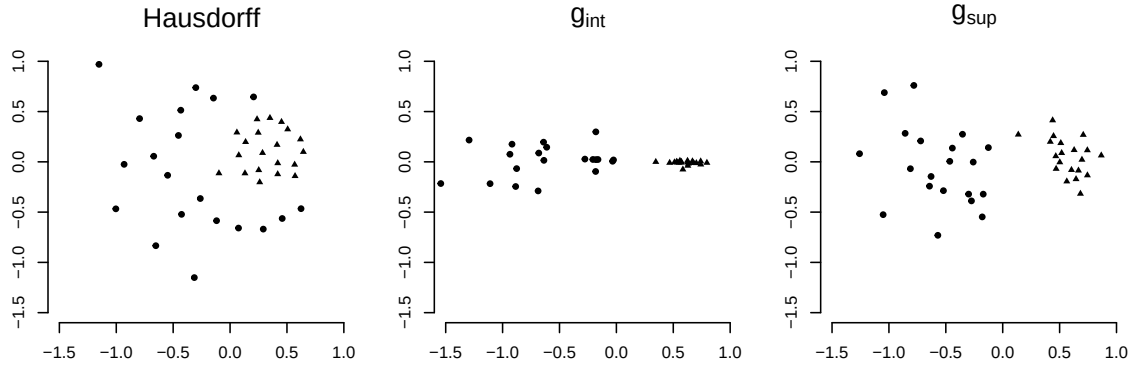

Figure S2.3: Binary classification $\Phi(\sigma)$ vs Π. Visualization of dissimilarities in a set of 40 patterns (20 patterns generated from $\Phi(\sigma)$ with $\sigma = 0.05$, denoted by circles, and 20 patterns from Π, denoted by triangles). The plots correspond to d_H (left), $\mathrm{d}_{int}[g]$ (middle) and $\mathrm{d}_{sup}[g]$ (right). The two-dimensional coordinates of the points of the plot are determined by the multidimensional scaling so that the distance of a pair of points in the plot is approximately proportional to the dissimilarity between the underlying pair of point patterns. This figure is the $2D$ counterpart of Fig. 4 from the main text.

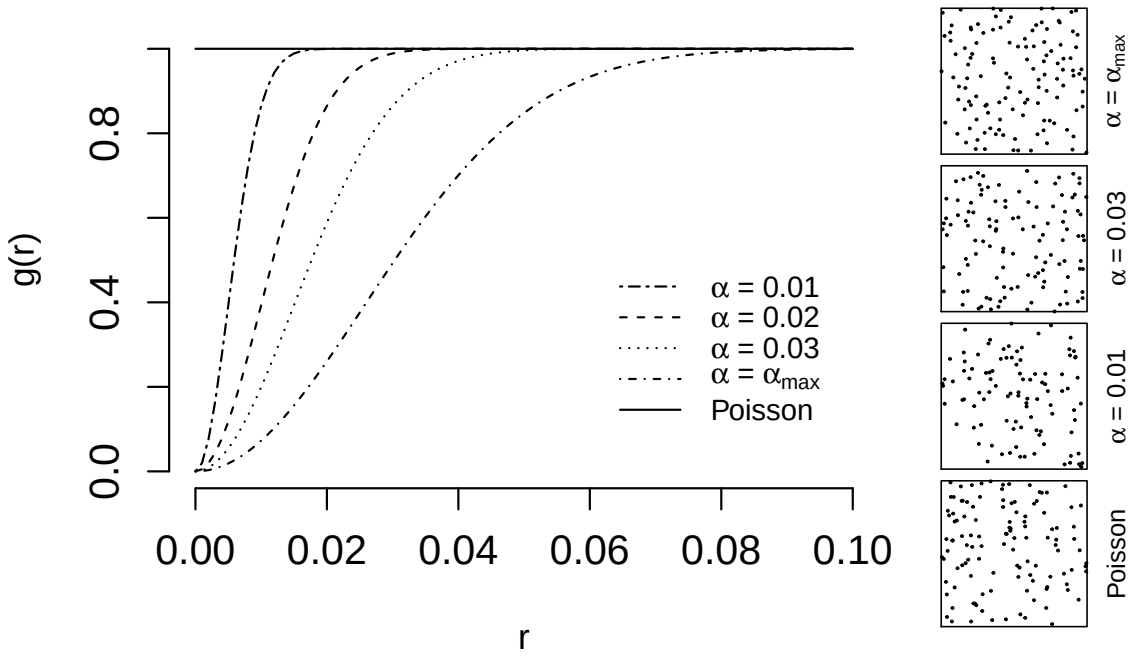

Figure S2.4: Illustration of the dependence of the behaviour of $\Psi(\alpha)$ on the model parameter α. For selected values of α, a realization of $\Psi(\alpha)$ is plotted as well as the evolution of the values of the pair correlation function g. The Poisson point process Π can be viewed as the limit of $\Psi(\alpha)$ for $\alpha \to 0$.

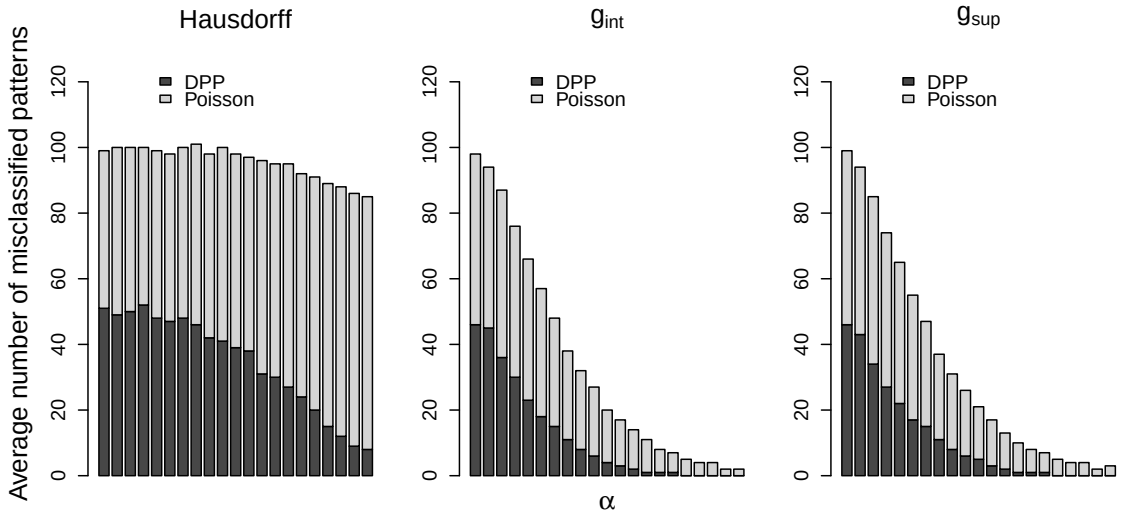

Figure S2.5: Binary classification $\Psi(\alpha)$ vs Π. For each α from the set $\mathcal{A} = \{0.0025, 0.0050, \ldots, 0.0500, \alpha_{max}\}$, the average number (from the 100 repetitions) of misclassified patterns in the testing data is plotted. We distinguish between realizations of the Gaussian determinantal point process $\Psi(\alpha)$ (dark grey) and the Poisson point process Π (light grey). The plots correspond to d_H (left), $\mathrm{d}_{int}[g]$ (middle) and $\mathrm{d}_{sup}[g]$ (right). For the smallest values of α (weak repulsive interactions), we observe the same number of misclassified realizations for both of the two groups regardless of the dissimilarity measure in use. Note that for bigger values of α, Π exhibits a higher variability in terms of realizations than $\Psi(\alpha)$. Thus, with growing α, the number of misclassified realizations of Π becomes greater than the number of misclassified realizations of $\Phi(\sigma)$. This phenomenon is more evident for d_H than for $\mathrm{d}_{int}[g]$ or $\mathrm{d}_{sup}[g]$ since the average number of misclassified patterns stays high for d_H, disregarding the value of α. For $\mathrm{d}_{int}[g]$ and $\mathrm{d}_{sup}[g]$, the average number of misclassified patterns goes to zero with α growing. For the largest values of α, only a few realizations of Π (on average) are misclassified.

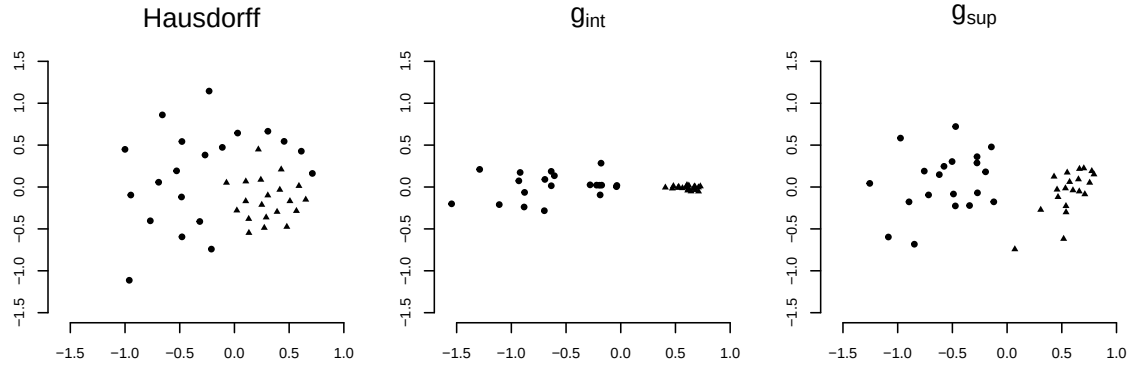

Figure S2.6: Binary classification $\Psi(\alpha)$ vs Π. Same situation as in Fig. S2.3, circles now correspond to the realizations of Π and triangles correspond to the realizations of $\Psi(\alpha)$ with $\alpha = \alpha_{max}$. Again, this figure is the $2D$ counterpart of Fig. 7 from the main text.

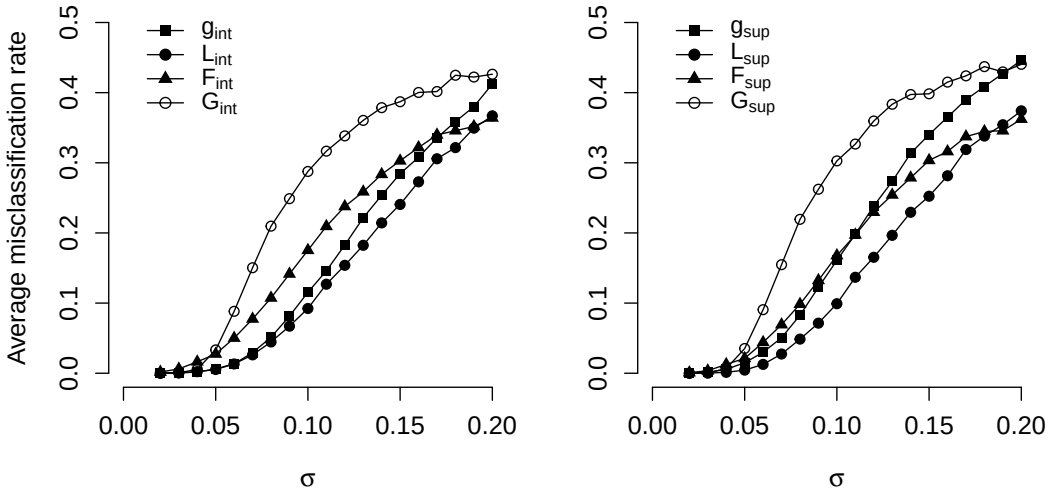

Figure S3.7: Average misclassification rates $\bar{\gamma}(\varphi_{f,int}[\sigma])$ (left) and $\bar{\gamma}(\varphi_{f,sup}[\sigma])$ (right) depending on the value of the model parameter σ and the choice of the functional summary characteristic $f \in \{g, L, F, G\}$.

S3 Experiment 1* – extension of Experiment 1 for L, F, G

This simulation experiment is an extension of Experiment 1 from the main text. We study the impact of the choice of the functional summary characteristic used to measure the dissimilarities.

Models, training and testing data Models Π, $\Phi(\sigma)$ and $\Psi(\alpha)$ as well as the training sets $\mathcal{T}(\sigma, 20, 20)$, $\mathcal{T}(\alpha, 20, 20)$ and the testing sets $\Gamma(\sigma, 100, 100)$, $\Gamma(\alpha, 100, 100)$ are the same as in Experiment 1 in the main text. To repeat, $\mathcal{T}(\sigma, 20, 20)$ resp. $\Gamma(\sigma, 100, 100)$ is composed from 20 resp. 100 realizations of $\Phi(\sigma)$ and 20 resp. 100 realizations of Π. The same holds for $\mathcal{T}(\alpha, 20, 20)$ and $\Gamma(\alpha, 100, 100)$, with $\Psi(\alpha)$ replacing $\Phi(\sigma)$.

Dissimilarity measures and classification scenarios Following the notation established in Experiment 1 in the main text, the dissimilarity measures $\mathrm{d}_{int}[f]$ and $\mathrm{d}_{sup}[f]$ are considered, $f \in \{L, F, G\}$, where L, F and G are the summary characteristics discussed in the opening section of the main text. For each $\sigma \in \Sigma$ and $f \in \{L, F, G\}$, we denote

$$\begin{aligned} \varphi_{f,int}[\sigma](\cdot) &= \varphi_{NW}(\,\cdot \mid \mathcal{T}(\sigma, 20, 20), \mathrm{d}_{int}[f], \mathrm{K}, h_{k_{LCV}}), \\ \varphi_{f,sup}[\sigma](\cdot) &= \varphi_{NW}(\,\cdot \mid \mathcal{T}(\sigma, 20, 20), \mathrm{d}_{sup}[f], \mathrm{K}, h_{k_{LCV}}), \end{aligned}$$

where K and $h_{k_{LCV}}$ are described in the main text, just before Experiment 1. The values of the average misclassification rates are compared with the reference values for $\varphi_{g,int}[\sigma]$ and $\varphi_{g,sup}[\sigma]$ from Experiment 1. For $\alpha \in \mathcal{A}$, the corresponding scenarios $\varphi_{f,int}[\alpha]$ and $\varphi_{f,sup}[\alpha]$ are considered.

Results for Thomas process The average misclassification rate is an increasing function of the model parameter σ, disregarding the choice of summary characteristics, see Fig. S3.7. It shows that $\varphi_{L,int}$ outperforms the other classifiers, although the difference between the average misclassification rate for $\varphi_{L,int}$ and $\varphi_{g,int}$ is, for small values of σ, negligible. The scenarios $\varphi_{F,int}$ and $\varphi_{G,int}$ lead to average misclassification rates above $\bar{\gamma}(\varphi_{g,int})$. Especially, $\bar{\gamma}(\varphi_{G,int})$

grows sharply towards the value 0.5. On the other hand, the difference between $\bar{\gamma}(\varphi_{F,int})$ and $\bar{\gamma}(\varphi_{g,int})$ is less prominent. Moreover, $\bar{\gamma}(\varphi_{F,int})$ is below $\bar{\gamma}(\varphi_{g,int})$ for the highest values of σ.

Similar observations apply to the maximum absolute deviation counterpart of the dissimilarity measures. Note that for g, the difference between the average misclassification rates for $\mathrm{d}_{g,int}$ and $\mathrm{d}_{g,sup}$ is the highest among the four summary characteristics. This confirms that $\mathrm{d}_{g,sup}$ suffers from the high variability of the estimator $g(r)$ for small values of r. Consequently, $\bar{\gamma}(\varphi_{F,int})$ is below $\bar{\gamma}(\varphi_{g,int})$ for $\sigma > 0.1$.

To show the variability of the individual misclassification rates, the 90% pointwise envelopes are plotted in Fig. S3.8 and Fig.S3.9. For numerical values of the average misclassification rate and the standard deviation for the sequence of the intermediate misclassification rates see Tab. S3.1 and Tab. S3.2.

	g_{int}		L_{int}		F_{int}		G_{int}	
σ	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0.02	< 0.001	0.001	0	0	0.002	0.003	0	0
0.03	< 0.001	0.001	0.001	0.002	0.002	0.007	< 0.001	0.001
0.04	0.002	0.003	0.002	0.004	0.016	0.011	0.006	0.006
0.05	0.006	0.008	0.005	0.006	0.027	0.017	0.033	0.012
0.06	0.013	0.013	0.013	0.009	0.050	0.024	0.088	0.022
0.07	0.029	0.014	0.026	0.013	0.077	0.027	0.150	0.029
0.08	0.051	0.020	0.045	0.018	0.107	0.033	0.210	0.035
0.09	0.081	0.028	0.067	0.021	0.142	0.033	0.249	0.033
0.10	0.115	0.029	0.092	0.022	0.175	0.039	0.288	0.034
0.11	0.146	0.033	0.127	0.025	0.209	0.037	0.317	0.041
0.12	0.183	0.034	0.154	0.027	0.238	0.038	0.338	0.050
0.13	0.222	0.037	0.182	0.032	0.259	0.039	0.360	0.043
0.14	0.254	0.040	0.214	0.033	0.284	0.038	0.379	0.045
0.15	0.284	0.042	0.241	0.040	0.303	0.042	0.387	0.043
0.16	0.308	0.045	0.273	0.044	0.322	0.047	0.400	0.053
0.17	0.336	0.044	0.306	0.045	0.340	0.048	0.402	0.042
0.18	0.359	0.046	0.322	0.041	0.346	0.046	0.425	0.048
0.19	0.380	0.046	0.349	0.050	0.352	0.048	0.422	0.050
0.20	0.412	0.049	0.367	0.044	0.364	0.048	0.426	0.050

Table S3.1: Binary classification $\Phi(\sigma)$ vs Π, $\sigma \in \Sigma$. For each σ, we report the average misclassification rate (Mean) and the standard deviation (SD) of the sequence of the individual misclassification rates for $\mathrm{d}_{int}[f]$, $f \in \{g, L, F, G\}$.

	g_{sup}		L_{sup}		F_{sup}		G_{sup}	
σ	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0.02	< 0.001	0.001	0	0	0.001	0.002	0	0
0.03	0.001	0.002	< 0.001	0.001	0.004	0.005	0.001	0.002
0.04	0.006	0.006	0.001	0.003	0.013	0.009	0.007	0.006
0.05	0.015	0.011	0.004	0.006	0.021	0.014	0.035	0.012
0.06	0.030	0.019	0.013	0.009	0.044	0.023	0.091	0.027
0.07	0.051	0.021	0.027	0.014	0.069	0.026	0.155	0.027
0.08	0.083	0.024	0.049	0.018	0.098	0.032	0.219	0.036
0.09	0.123	0.030	0.071	0.021	0.132	0.034	0.262	0.032
0.10	0.161	0.032	0.099	0.024	0.168	0.041	0.303	0.035
0.11	0.198	0.035	0.137	0.026	0.197	0.036	0.327	0.039
0.12	0.239	0.038	0.165	0.028	0.230	0.042	0.360	0.047
0.13	0.273	0.045	0.197	0.033	0.254	0.040	0.383	0.041
0.14	0.314	0.047	0.229	0.037	0.279	0.043	0.397	0.040
0.15	0.340	0.043	0.252	0.035	0.303	0.037	0.398	0.051
0.16	0.365	0.045	0.282	0.043	0.316	0.042	0.415	0.050
0.17	0.390	0.043	0.319	0.039	0.337	0.047	0.424	0.042
0.18	0.409	0.042	0.338	0.049	0.345	0.048	0.437	0.043
0.19	0.427	0.042	0.354	0.042	0.346	0.049	0.430	0.054
0.20	0.446	0.042	0.374	0.044	0.362	0.052	0.440	0.047

Table S3.2: Binary classification $\Phi(\sigma)$ vs Π, $\sigma \in \Sigma$. This table is a counterpart to Tab. S3.1 for $\mathrm{d}_{sup}[f]$, $f \in \{g, L, F, G\}$.

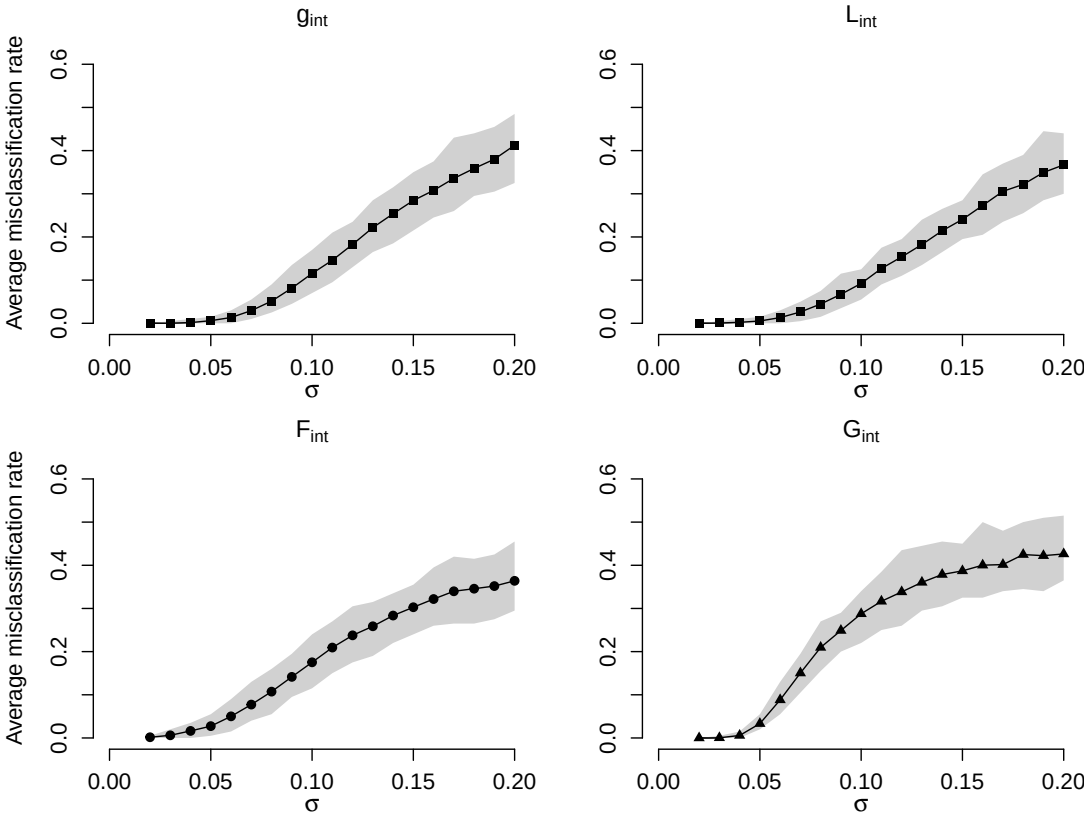

Figure S3.8: Binary classification $\Phi(\sigma)$ vs Π, $\sigma \in \Sigma$. Average misclassification rate and the 90% pointwise envelope for the sequence of the individual misclassification rates are plotted for $\mathrm{d}_{int}[g]$ (top-left), $\mathrm{d}_{int}[L]$ (top-right), $\mathrm{d}_{int}[F]$ (bottom-left) and $\mathrm{d}_{int}[G]$ (bottom-right). For small values of σ, we start with a very narrow envelope whose width is growing with σ going to 0.20.

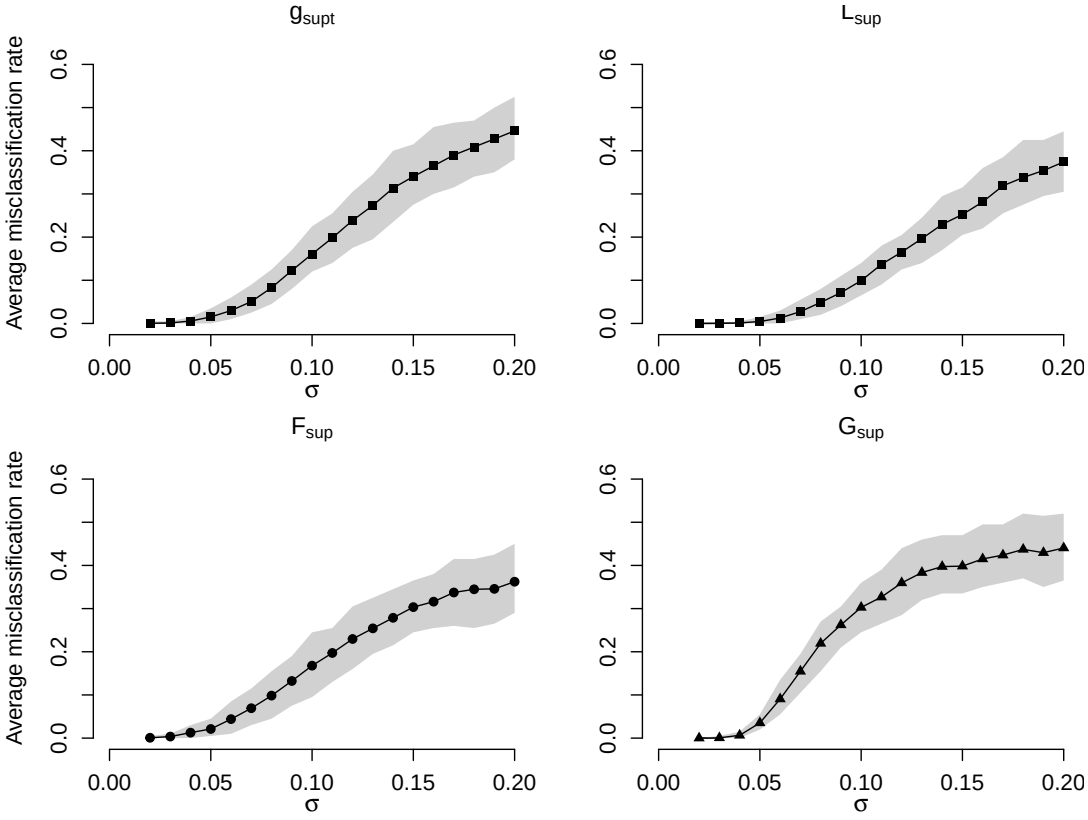

Figure S3.9: Binary classification $\Phi(\sigma)$ vs Π, $\sigma \in \Sigma$. This figure is a counterpart of Fig. S3.8 for $\mathrm{d}_{sup}[f]$, $f \in \{g, L, F, G\}$. Differences between the results for the integral dissimilarity measure and the supremum dissimilarity measure are rather negligible.

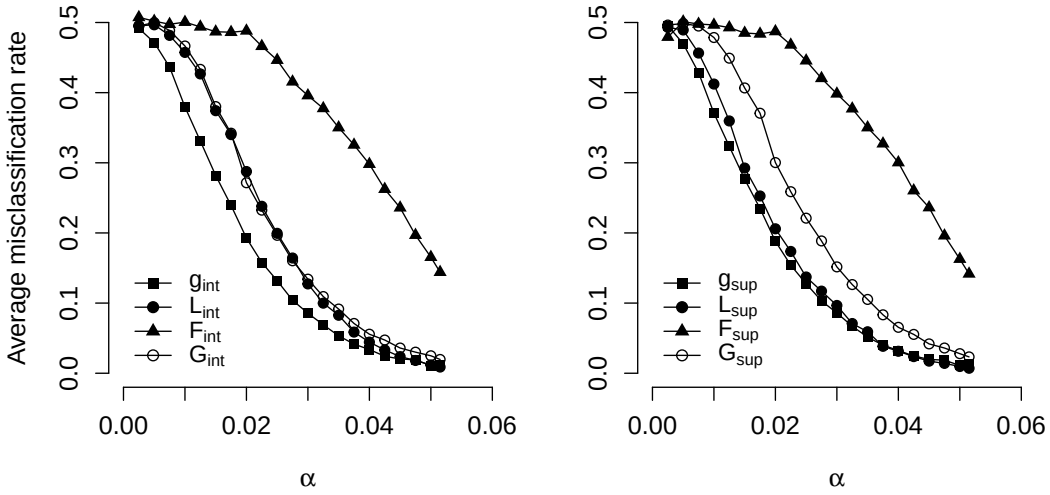

Figure S3.10: Average misclassification rates $\bar{\gamma}(\varphi_{f,int}[\alpha])$ (left) and $\bar{\gamma}(\varphi_{f,sup}[\alpha])$ (right) depending on the value of the model parameter α and the choice of the functional summary characteristic $f \in \{g, L, F, G\}$.

Results for Gaussian determinantal process In this case, the average misclassification rate is a decreasing function of α, regardless of the choice of the summary characteristic f, see Fig. S3.10. It shows that $\varphi_{g,int}$ outperforms the other classifiers $\varphi_{\cdot,int}$. The choice of L and G results in average misclassification rates of negligible difference. The classification scenario $\varphi_{F,int}$ results in the highest average misclassification rate, for all considered values of α.

Similar observations apply to the maximum absolute deviation counterparts of the dissimilarity measures. There is a difference between the average misclassification rates for $\varphi_{L,int}$ and $\varphi_{L,sup}$ in favor of the latter, and $\bar{\gamma}(\varphi_{L,sup})$ copy the values of $\bar{\gamma}(\varphi_{g,sup})$ for all $\alpha \in \mathcal{A}$. The difference between $\bar{\gamma}(\varphi_{L,sup})$ and $\bar{\gamma}(\varphi_{G,sup})$ is much larger than for the integral scenarios.

The 90% pointwise envelopes are plotted in Fig. S3.11 for the integral dissimilarity measures and in Fig.S3.12 for their maximum absolute deviation counterparts. Numerical values of the average misclassification rate and the standard deviation for the sequence of the individual misclassification rates are summarized in Tab. S3.3 and Tab. S3.4.

Summary This experiment indicates that choosing the appropriate dissimilarity measure (the appropriate functional summary characteristic) is crucial. For the binary classification of aggregation/repulsive interactions vs complete spatial randomness, the second-order characteristics g and L result in the most satisfying performance of the proposed Bayes classifier, combined with the k-nearest neighbors algorithms and the kernel regression method.

	g_{int}		L_{int}		F_{int}		G_{int}	
α	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0.0025	0.492	0.037	0.495	0.031	0.507	0.122	0.495	0.039
0.0050	0.471	0.042	0.497	0.035	0.502	0.032	0.500	0.035
0.0075	0.437	0.039	0.482	0.036	0.498	0.035	0.489	0.034
0.0100	0.380	0.050	0.457	0.046	0.501	0.034	0.467	0.048
0.0125	0.331	0.045	0.427	0.041	0.494	0.035	0.433	0.045
0.0150	0.282	0.040	0.374	0.047	0.487	0.037	0.380	0.048
0.0175	0.240	0.034	0.340	0.051	0.486	0.041	0.342	0.047
0.0200	0.193	0.032	0.288	0.049	0.488	0.040	0.272	0.037
0.0225	0.158	0.034	0.238	0.051	0.466	0.050	0.233	0.035
0.0250	0.132	0.027	0.199	0.044	0.446	0.049	0.197	0.037
0.0275	0.104	0.023	0.164	0.038	0.416	0.048	0.160	0.031
0.0300	0.086	0.022	0.127	0.036	0.396	0.052	0.134	0.025
0.0325	0.069	0.021	0.100	0.034	0.378	0.043	0.109	0.025
0.0350	0.053	0.019	0.082	0.025	0.350	0.041	0.092	0.022
0.0375	0.042	0.016	0.059	0.021	0.326	0.049	0.071	0.021
0.0400	0.034	0.017	0.044	0.020	0.298	0.043	0.056	0.019
0.0425	0.024	0.013	0.033	0.016	0.262	0.036	0.047	0.019
0.0450	0.020	0.014	0.024	0.014	0.236	0.033	0.036	0.016
0.0475	0.019	0.012	0.018	0.012	0.197	0.032	0.030	0.014
0.0500	0.011	0.008	0.012	0.008	0.165	0.032	0.025	0.012
α_{max}	0.012	0.009	0.001	0.007	0.144	0.032	0.020	0.010

Table S3.3: Binary classification $\Psi(\alpha)$ vs Π, $\alpha \in \mathcal{A}$. For each α, we report the average misclassification rate (Mean) and the standard deviation (SD) of the sequence of the individual misclassification rates for $\mathrm{d}_{int}[f],\ f \in \{g, L, F, G\}$.

	g_{sup}		L_{sup}		F_{sup}		G_{sup}	
α	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0.0025	0.496	0.033	0.493	0.031	0.479	0.116	0.496	0.037
0.0050	0.470	0.046	0.489	0.037	0.501	0.034	0.499	0.038
0.0075	0.429	0.037	0.456	0.041	0.498	0.036	0.495	0.039
0.0100	0.371	0.047	0.412	0.052	0.497	0.040	0.479	0.039
0.0125	0.323	0.041	0.360	0.045	0.493	0.034	0.449	0.041
0.0150	0.277	0.037	0.293	0.035	0.485	0.040	0.407	0.044
0.0175	0.235	0.035	0.253	0.035	0.484	0.043	0.371	0.043
0.0200	0.188	0.030	0.206	0.036	0.487	0.041	0.300	0.040
0.0225	0.154	0.029	0.174	0.035	0.468	0.050	0.259	0.035
0.0250	0.128	0.025	0.137	0.026	0.446	0.050	0.221	0.043
0.0275	0.103	0.024	0.117	0.027	0.420	0.050	0.189	0.032
0.0300	0.086	0.023	0.097	0.025	0.398	0.051	0.152	0.026
0.0325	0.067	0.020	0.071	0.019	0.377	0.045	0.126	0.027
0.0350	0.051	0.017	0.059	0.019	0.350	0.039	0.105	0.024
0.0375	0.040	0.015	0.038	0.012	0.327	0.047	0.083	0.021
0.0400	0.032	0.016	0.031	0.013	0.300	0.042	0.065	0.020
0.0425	0.025	0.014	0.024	0.012	0.260	0.033	0.055	0.019
0.0450	0.020	0.013	0.017	0.008	0.236	0.033	0.042	0.016
0.0475	0.019	0.012	0.014	0.009	0.196	0.031	0.036	0.015
0.0500	0.011	0.007	0.010	0.007	0.163	0.030	0.028	0.013
α_{max}	0.013	0.009	0.007	0.006	0.142	0.034	0.023	0.010

Table S3.4: Binary classification $\Psi(\alpha)$ vs Π, $\alpha \in \mathcal{A}$. This table is a counterpart of Tab. S3.3 for $\mathrm{d}_{sup}[f]$, $f \in \{g, L, F, G\}$.

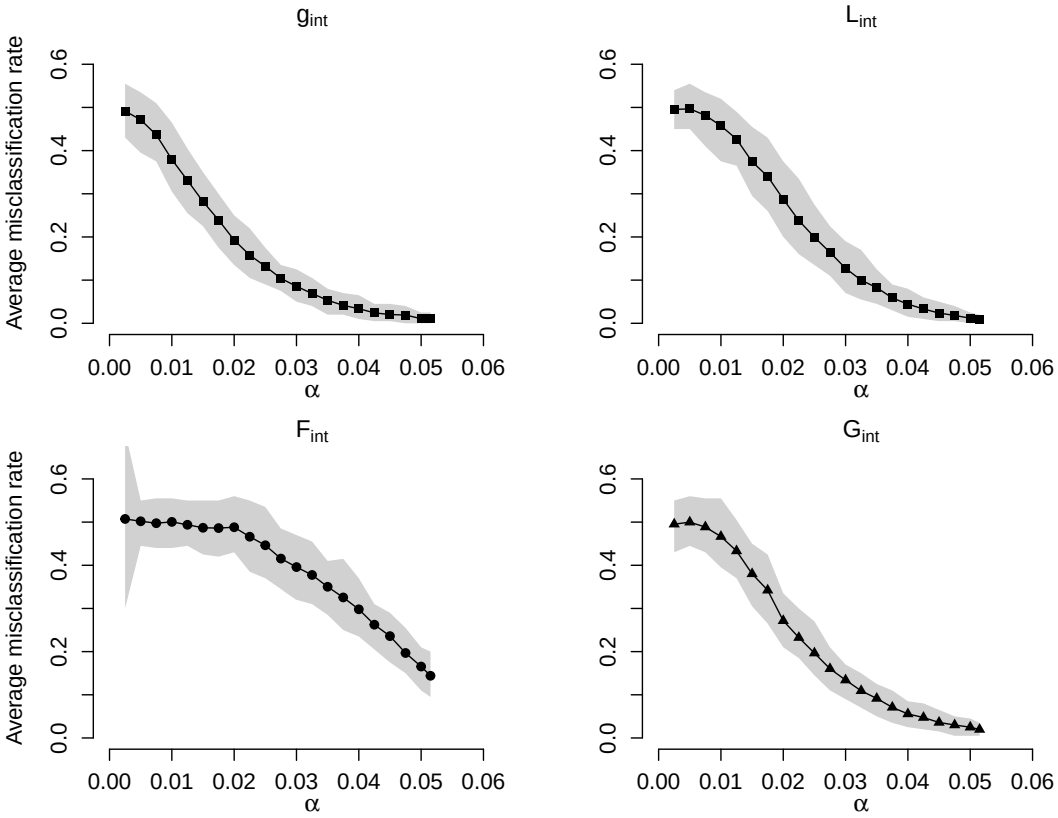

Figure S3.11: Binary classification $\Psi(\alpha)$ vs Π, $\alpha \in \mathcal{A}$. Average misclassification rate and the 90% pointwise envelope for the sequence of the individual misclassification rates are plotted for $\mathrm{d}_{int}[g]$ (top-left), $\mathrm{d}_{int}[L]$ (top-right), $\mathrm{d}_{int}[F]$ (bottom-left) and $\mathrm{d}_{int}[G]$ (bottom-right). For small values of α, we start with a wide envelope whose width is decreasing with α going to α_{max}.

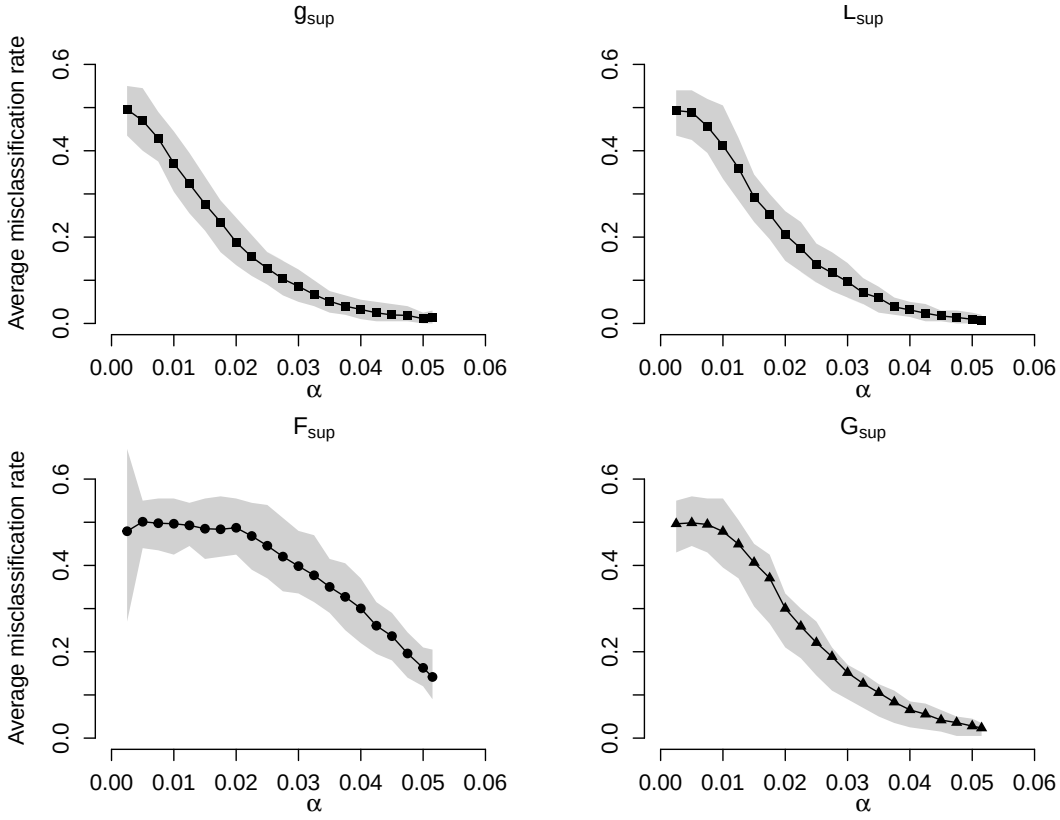

Figure S3.12: Binary classification $\Psi(\alpha)$ vs Π, $\alpha \in \mathcal{A}$. This figure is a counterpart of Fig. S3.11 for $\mathrm{d}_{sup}[f]$, $f \in \{g, L, F, G\}$. Differences between the results for the integral dissimilarity measures and their maximum absolute deviation counterparts are negligible.

S4 Impact of the choice of the kernel function

In this experiment, we study the impact of the choice of the kernel function and the automatic procedure to choose the value of the parameter k on the performance of the proposed classifiers.

Models, training and testing data We consider the classification schemes $\Phi(\sigma)$ vs Π, $\sigma \in \{0.05, 0.10, 0.15\}$, and $\Psi(\alpha)$ vs Π, with $\alpha \in \{0.02, 0.03, \alpha_{max}\}$, respectively. The training set $\mathcal{T}(\sigma) = \mathcal{T}(\sigma, 20, 20)$ consists of 20 realizations of $\Phi(\sigma)$ and 20 realizations of Π. The testing set $\Gamma(\sigma) = \Gamma(\sigma, 100, 100)$ is composed of 100 realizations of $\Phi(\sigma)$ and 100 realizations of Π. Replacing $\Phi(\sigma)$ with $\Psi(\alpha)$, we obtain the training set $\mathcal{T}(\alpha)$ and the testing set $\Gamma(\alpha)$.

Dissimilarity measures and classification scenarios We use $\mathrm{d}_{int}[g]$ as the dissimilarity measure. For $\sigma \in \{0.05, 0.10, 0.15\}$ we denote

$$\begin{aligned} \varphi_{NN,\sigma}[glob](\cdot) &= \varphi_{NW}(\,\cdot \mid \mathcal{T}(\sigma), \mathrm{d}_{int}[g], \mathrm{K}_{unif}, h_{k_{GCV}}), \\ \varphi_{NW,\sigma}[loc](\cdot) &= \varphi_{NW}(\,\cdot \mid \mathcal{T}(\sigma), \mathrm{d}_{int}[g], \mathrm{K}_{E}, h_{k_{LCV}}), \end{aligned}$$

where K_{unif} is the uniform kernel, K_E is the Epanechnikov kernel, and $h_{k_{GCV}}$ resp. $h_{k_{LCV}}$ denotes the bandwidth corresponding to the k-nearest neighbours selection method with automatic global resp. local procedure of the choice of parameter k. More details can be found in the main text within the section devoted to the description of the supervised classification task. For $\alpha_1 \in \{0.02, 0.03, \alpha_{max}\}$, the corresponding scenarios $\varphi_{NN,\alpha}[glob]$ and $\varphi_{NW,\alpha}[loc]$ are considered.

Results First, the kernel regression method combined with the uniform kernel and k-nearest neighbours procedure for the choice of h corresponds to the k-nearest neighbour classifier. With the global choice of k, the leave-one-out cross-validation is used to find one (global) optimal value of k that is used for all patterns in the testing set. For the local choice of k, we choose a different optimal k for every element in the testing set. For both situations, $\Phi(\sigma)$ vs Π and $\Psi(\alpha)$ vs Π, the differences between the average misclassification rates $\varphi_{NN}[glob]$ and $\varphi_{NW}[loc]$ are almost indistinguishable, with a small favor for the latter. This holds for all values of σ resp. α, see Tab. S4.5 resp. Tab. S4.6. It appears that in our setting, the choice of the kernel function is not crucial. The local choice of k can have a positive impact on the performance of the classifier. However, it does not make a big difference. In contrast, we have shown that the choice of the dissimilarity measure is essential, see Experiment 1 in the main text and its extension in Sect. S3.

	Mean			**SD**		
σ	0.05	0.10	0.15	0.05	0.10	0.15
$\varphi_{NN,\sigma}[glob]$	0.011	0.119	0.288	0.010	0.032	0.039
$\varphi_{NW,\sigma}[loc]$	0.006	0.110	0.284	0.007	0.026	0.039

Table S4.5: Binary classification $\Phi(\sigma)$ vs Π, $\sigma \in \{0.05, 0.10, 0.15\}$. For each σ, the average misclassification rates (Mean) and the standard deviation of the sequence of individual misclassification rates are reported for the two scenarios $\varphi_{NN,\sigma}[glob]$ and $\varphi_{NW,\sigma}[loc]$.

	Mean			SD		
α	0.02	0.03	α_{max}	0.02	0.03	α_{max}
$\varphi_{NN,\alpha}[glob]$	0.189	0.090	0.014	0.028	0.024	0.010
$\varphi_{NW,\alpha}[loc]$	0.189	0.086	0.011	0.029	0.023	0.009

Table S4.6: Binary classification $\Psi(\alpha)$ vs Π, $\alpha \in \{0.02, 0.03, \alpha_{max}\}$. For each α, the average misclassification rates (Mean) and the standard deviation of the sequence of individual misclassification rates are reported for the two scenarios $\varphi_{NN,\alpha}[glob]$ and $\varphi_{NW,\alpha}[loc]$.

S5 Impact of the choice of the constant R

This experiment studies the impact of the choice of the constant R on the proposed classifiers. This constant appears in the formulas for dissimilarity measures based on point process functional summary characteristics, see equation (4) in the main text. We consider the underlying dissimilarity measure to be $\mathrm{d}_{int}[g]$, which performed the best in Experiment 1 in the main text. The maximum absolute deviation counterpart of the dissimilarity measure is not considered here since, in the models in question, the supremum is almost exclusively attained in the neighbourhood of 0, making the influence of R negligible. The values of the model parameters σ and α are fixed in this experiment.

Models, training and testing data Using the notation from Experiment 1, we fix $\sigma = 0.1$ and $\alpha = 0.03$, and the realizations from $\Phi(\sigma)$ or $\Psi(\alpha)$ are compared with the realizations of the Poisson point process Π. Training data are $\mathcal{T}(\sigma, 20, 20)$ and $\mathcal{T}(\alpha, 20, 20)$, respectively. The testing data are $\Gamma(\sigma, 100, 100)$ and $\Gamma(\alpha, 100, 100)$, respectively.

Dissimilarity measure and classification scenarios To stress the dependence on R, we suppress the subscript and the dependence on g and write $\mathrm{d}[R](\cdot,\cdot) = \mathrm{d}_{int}(g,\cdot,\cdot)$. Let $\mathcal{R} = \{0.01, 0.02, \ldots, 0.25\}$, this choice is related to the observation window being the unit square. For each $R \in \mathcal{R}$, we consider the classification scenarios

$$\begin{aligned} \varphi_R[\sigma](\cdot) &= \varphi_{NW}(\,\cdot \mid \mathcal{T}(\sigma, 20, 20), \mathrm{d}[R], \mathrm{K}, h_{k_{LCV}}), \\ \varphi_R[\alpha](\cdot) &= \varphi_{NW}(\,\cdot \mid \mathcal{T}(\alpha, 20, 20), \mathrm{d}[R], \mathrm{K}, h_{k_{LCV}}). \end{aligned}$$

We report the average misclassification rates $\bar{\gamma}(\varphi_R[\sigma])$ and $\bar{\gamma}(\varphi_R[\alpha])$ as functions of the argument R.

Results for Thomas process Fig. S5.13 shows high average misclassification rates $\bar{\gamma}(\varphi_R[\sigma])$ for small values of R. In this case, the range for the argument r in the definition of d_{int} is too short to capture enough information about the examined point patterns. Also, the estimator of $g(r)$ has a high variance for r close to 0. With increasing R, $\bar{\gamma}(\varphi_R[\sigma])$ steeply decreases to 0.1. The minimum value is attained for $R = 0.15$. For large values of r, the estimator of $g(r)$ again has higher variance, since it is based on fewer pairs of points with higher weights. This causes $\bar{\gamma}(\varphi_R[\sigma)$ to increase slightly for $R > 0.15$.

Results for Gaussian determinantal process According to Fig. S5.13, the average misclassification rate $\bar{\gamma}(\varphi_R[\alpha])$ is approximately 0.08 for all values of R. In fact, it starts at 0.085 for $R = 0.01$, decreases to 0.072 for $R = 0.03$ and then increases slightly to 0.086 for $R = 0.25$.

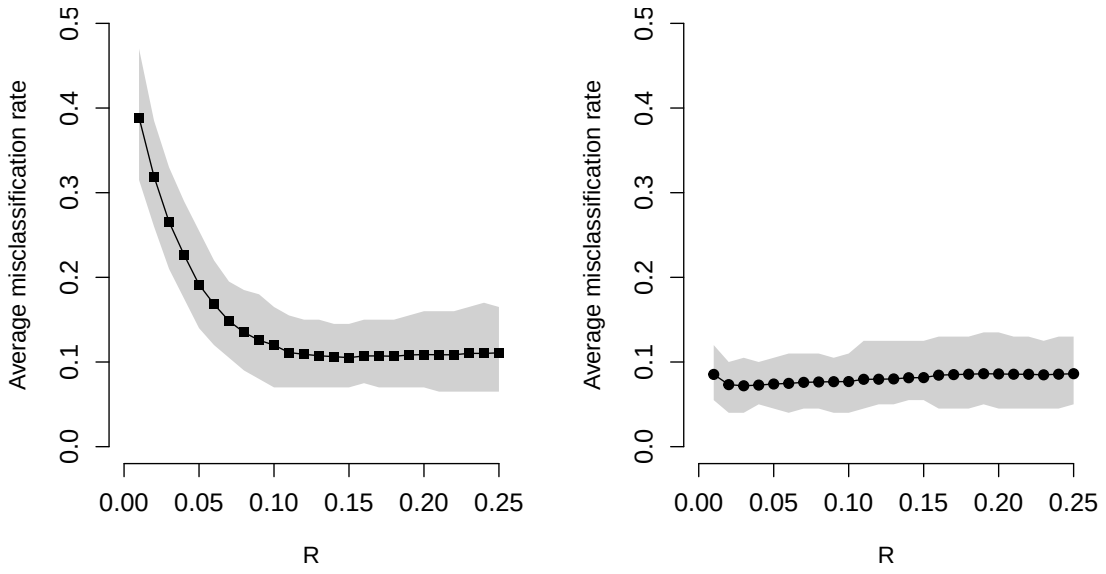

Figure S5.13: Average misclassification rates $\bar{\gamma}(\varphi_R[\sigma])$ with $\sigma = 0.1$ (left) and $\bar{\gamma}(\varphi_R[\alpha])$ with $\alpha = 0.03$ (right) are plotted as functions of the parameter R. To illustrate the variability of the individual misclassification rates, pointwise 90% envelopes are plotted.

Summary In the classification of $\Phi(\sigma)$ vs Π, inappropriate choice of R has a negative impact on the performance of $\varphi_R[\sigma]$. Preliminary experiments should be performed to find an appropriate value of R, depending on the size and shape of the observation window. Moreover, choosing a smaller R can reduce computational costs (the bottleneck is the estimation of g for the computation of the dissimilarities). On the other hand, the performance of the classifier in the $\Psi(\alpha)$ vs Π setting is robust with respect to the choice of R.

S6 Impact of size and composition of training data

This experiment studies the effect of the size and composition of the training data on the performance of the proposed classifiers. We extend the results of Experiment 1 in the main text by considering six different choices of the training set.

Models, training and testing data Take $\sigma \in \{0.05, 0.1, 0.15\}$ and $\alpha \in \{0.02, 0.03, \alpha_{max}\}$. We focus on the binary classification $\Phi(\sigma)$ vs Π and $\Psi(\alpha)$ vs Π, respectively. Let $n_1, n_2 \in \mathbb{N}$. The training set $\mathcal{T}(\sigma, n_1, n_2)$ consists of n_1 realizations of $\Phi(\sigma)$ and n_2 realizations of Π. The testing set $\Gamma(\sigma) = \Gamma(\sigma, 100, 100)$ is the same as in Experiment 1 in the main text. Replacing $\Phi(\sigma)$ with $\Psi(\alpha)$, we obtain the training set $\mathcal{T}(\alpha, n_1, n_2)$ and the testing set $\Gamma(\alpha)$.

Dissimilarity measures and classification scenarios We choose $\mathrm{d}_{int}[g]$ and $\mathrm{d}_{sup}[g]$ to be the dissimilarity measures. For $\sigma \in \{0.05, 0.1, 0.15\}$ and $n_1, n_2 \in \mathbb{N}$ we denote by $\varphi_{\sigma,int}[n_1, n_2]$ the classification rule

$$\varphi_{NW}(\,\cdot \mid \mathcal{T}(\sigma, n_1, n_2), \mathrm{d}_{int}[g], \mathrm{K}, h_{k_{LCV}}),$$

and by $\varphi_{\sigma,sup}[n_1, n_2]$ the rule

$$\varphi_{NW}(\,\cdot \mid \mathcal{T}(\sigma, n_1, n_2), \mathrm{d}_{sup}[g], \mathrm{K}, h_{k_{LCV}}),$$

where K and $h_{k_{LCV}}$ are as in Experiment 1 in the main text. In the notation, we suppress the dependence of the classification rule φ on the functional summary statistic. The average misclassification rates $\bar{\gamma}\left(\varphi_{\sigma,int}[n_1, n_2]\right)$ and $\bar{\gamma}\left(\varphi_{\sigma,sup}[n_1, n_2]\right)$ will be compared with $\bar{\gamma}\left(\varphi_{\sigma,int}[20, 20]\right)$ and $\bar{\gamma}\left(\varphi_{\sigma,sup}[20, 20]\right)$ from Experiment 1 in the main text. For the parameter $\alpha \in \{0.02, 0.03, \alpha_{max}\}$, the corresponding scenarios $\varphi_{\alpha,int}[n_1, n_2]$ and $\varphi_{\alpha,sup}[n_1, n_2]$ are considered.

Results for Thomas process In the upper part of Table S6.7, we set $n_1 = n_2$ and increase the value of n_1. The training set is thus composed of the same amount of realizations from both models. We expect the average misclassification rates to decrease with increasing n_1. In fact, this is observed in the upper part of Table S6.7, the exception being the integral dissimilarity measure with $\sigma = 0.15$, where the trend is not so clear.

For $\sigma = 0.05$, i.e. strong attractive interactions, the average misclassification rates corresponding to the unbalanced training data (bottom part of Tab S6.7) are nearly the same as the values for the balanced training data of the same total size ($n_1 = n_2 = 80$). For $\sigma = 0.10$, the average misclassification rates for unbalanced training data are higher than those corresponding to $n_1 = n_2 = 20$; for $\sigma = 0.15$ they are higher than all values corresponding to balanced training data.

In Experimen 1 in the main text, we have observed that the realizations of $\Phi(\sigma)$ tend to be misclassified more often than the realization of Π, even though this phenomenon is less evident for large values of σ. In the unbalanced situation $n_1 = 120, n_2 = 40$, the small number of realizations of Π in the training set results in incorrect labeling of realizations of Π in the testing set. On the other hand, having only 40 realizations of $\Phi(\sigma)$ in the training set causes an even higher number of misclassified realizations of $\Phi(\sigma)$. Having in mind that, roughly speaking, it is easier to distinguish the realizations of $\Phi(\sigma)$ from those of Π for $\sigma = 0.10$ than for $\sigma = 0.15$, it is not surprising that for $\sigma = 0.10$ situation $n_1 = 40, n_2 = 120$ leads to a higher average misclassification rate than situation $n_1 = 120, n_2 = 40$ and vice versa for $\sigma = 0.15$.

Results for Gaussian determinantal process Similar conclusions can be drawn for the classification $\Psi(\alpha)$ vs Π. For $\alpha = 0.02$, the lowest average misclassification rate is obtained for $n_1 = n_2 = 20$, for both integral and supremum dissimilarity measures (while noting that the average misclassification rates are comparable in all scenarios with balanced training data). For $\alpha = 0.03$, the lowest values are obtained for $n_1 = n_2 = 80$. Similarly for $\alpha = \alpha_{max}$, where only the case with $n_1 = 40, n_2 = 120$ resulted in a marginally lower average misclassification rate.

In general, Table S6.8 confirms the observation that for the balanced training set, the average misclassification rate decreases with increasing n_1 and that satisfactory results can be obtained with a relatively small value of n_1. For the unbalanced training set, we can make analogous observations as for the Thomas process. Note that here, Π is, loosely speaking, the model with higher variability among realizations, and thus realizations of Π tend to be misclassified more often. Therefore, the role of the situation $n_1 = 120, n_2 = 40$ and $n_1 = 40, n_2 = 20$ is interchanged compared to the classification $\Phi(\sigma)$ vs Π.

Summary For both situations, Thomas vs Poisson and Poisson vs Gaussian DPP, we have observed that the average misclassification rate can be reduced by increasing the size of the training set, provided that the sample is balanced. However, the classification is already satisfactory for moderate sample sizes, say $n_1 = n_2 = 20$. For weak and moderate interactions (larger values of σ, smaller values of α), using unbalanced training data leads to higher average

	Training		$\mathrm{d}_{int}[g]$			$\mathrm{d}_{sup}[g]$		
σ	n_1	n_2	0.05	0.10	0.15	0.05	0.10	0.15
	10	10	0.006	0.133	0.314	0.016	0.180	0.376
	20	20	0.006	0.115	0.284	0.015	0.161	0.340
	40	40	0.003	0.108	0.296	0.007	0.139	0.332
	80	80	0.002	0.106	0.282	0.006	0.133	0.310
	120	40	0.002	0.128	0.384	0.008	0.177	0.411
	40	120	0.004	0.148	0.348	0.008	0.185	0.390

Table S6.7: Binary classification $\Phi(\sigma)$ vs Π, $\sigma \in \{0.05, 0.10, 0.15\}$. In each row, we report the numbers of patterns per class n_1 and n_2 and the average misclassification rates $\bar{\gamma}(\varphi_{\sigma,int}[n_1, n_2])$ and $\bar{\gamma}(\varphi_{\sigma,sup}[n_1, n_2])$, respectively.

	Training		$\mathrm{d}_{int}[g]$			$\mathrm{d}_{sup}[g]$		
α	n_1	n_2	0.02	0.03	α_{max}	0.02	0.03	α_{max}
	10	10	0.215	0.107	0.015	0.214	0.102	0.016
	20	20	0.193	0.086	0.012	0.188	0.086	0.013
	40	40	0.201	0.096	0.008	0.200	0.093	0.010
	80	80	0.199	0.081	0.007	0.195	0.082	0.008
	120	40	0.254	0.127	0.014	0.242	0.120	0.015
	40	120	0.268	0.100	0.006	0.262	0.101	0.007

Table S6.8: Binary classification $\Psi(\alpha)$ vs Π, $\alpha \in \{0.02, 0.03, \alpha_{max}\}$. In each row, we report the number of patterns per class n_1 and n_2 and the average misclassification rates $\bar{\gamma}(\varphi_{\alpha,int}[n_1, n_2])$ and $\bar{\gamma}(\varphi_{\alpha,sup}[n_1, n_2])$, respectively.

misclassification rates than using balanced training data of smaller total size (say, $n_1 = 20$). These observations hold for both the dissimilarity measures $\mathrm{d}_{int}[g]$ and $\mathrm{d}_{sup}[g]$.

S7 Experiment 3

This section provides additional visual material for Experiment 3 from the main text. A binary classification is considered, with the two models at hand that belong to the same parametric family and differ in the value of the model parameter.

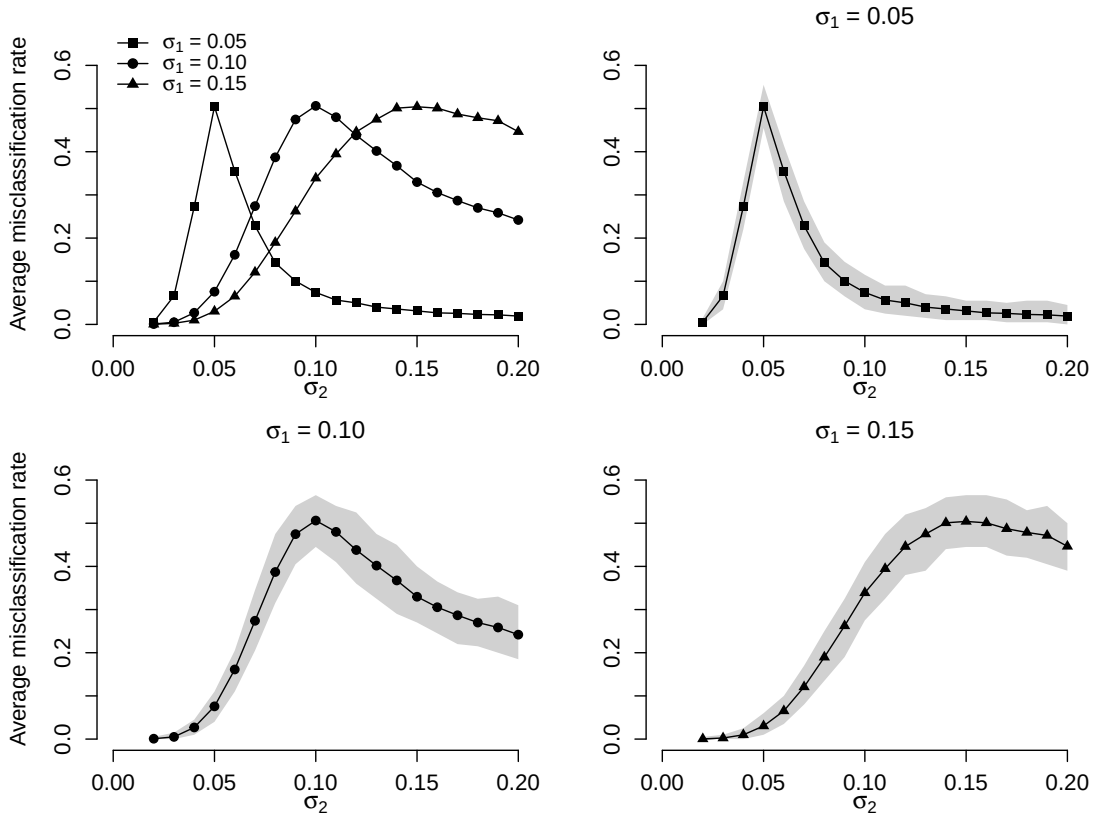

Figure S7.14: Binary classification $\Phi(\sigma_1)$ vs $\Phi(\sigma_2)$, $\sigma_1 \in \{0.05, 0.10, 0.15\}$, $\sigma_2 \in \Sigma$. All realizations are observed on the unit square. The average misclassification rates for $\varphi_{g,sup}[\sigma_1, \sigma_2]$ are plotted together with the 90% pointwise envelope for the sequence of the individual misclassification rates. This figure is a counterpart of Fig. 9 in the main text.

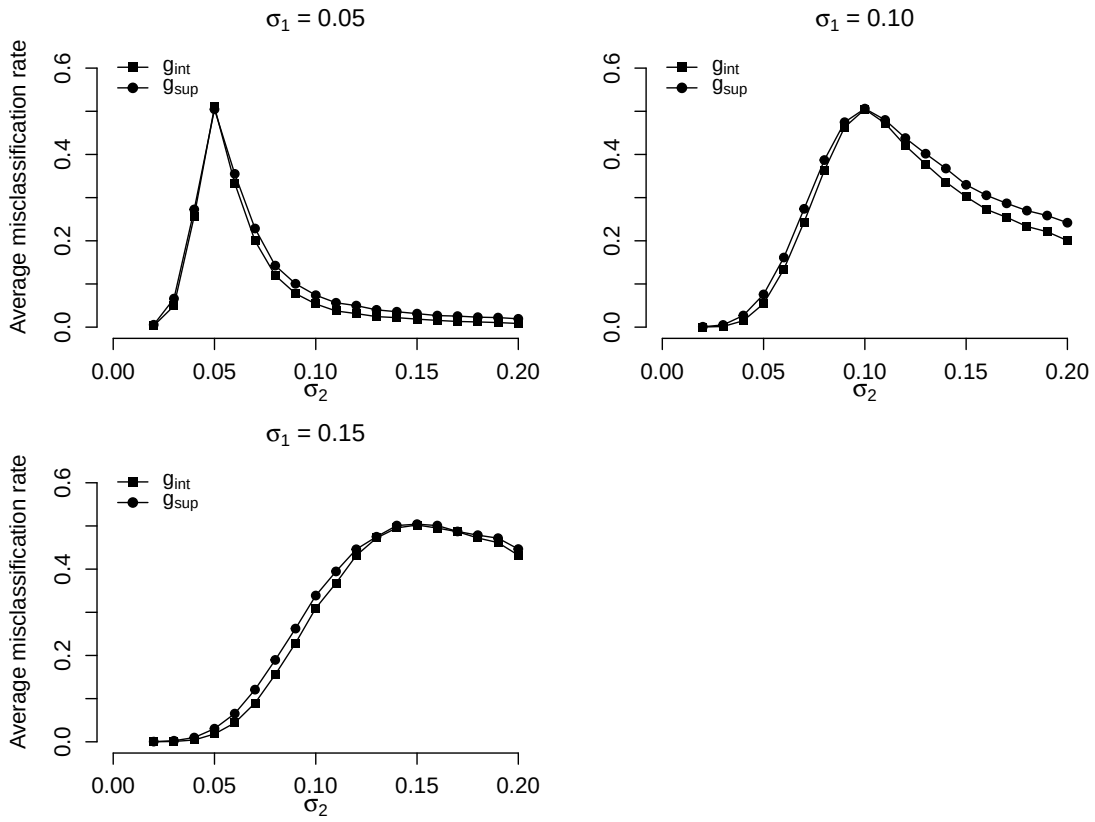

Figure S7.15: Binary classification $\Phi(\sigma_1)$ vs $\Phi(\sigma_2)$, $\sigma_1 \in \{0.05, 0.10, 0.15\}$, $\sigma_2 \in \Sigma$. All realizations are observed on the unit square. To compare the scenarios with $\mathrm{d}_{int}[g]$ resp. $\mathrm{d}_{sup}[g]$, the average misclassification rates for both of the classification scenarios are plotted. The plots correspond to the different values of σ_1: $\sigma_1 = 0.05$ (top-left), $\sigma_1 = 0.10$ (top-right) and $\sigma_1 = 0.15$ (bottom-left).

S8 Experiment 4* – extension of Experiment 4 for a larger observation window

In this section, we repeat the situation from Experiment 3 in the main text with point patterns simulated in a larger observation window. Thus, a larger number of points is observed in each pattern, implying a more precise estimation of the pair correlation function.

Models, training and testing data Again we consider binary classification $\Phi(\sigma_1)$ vs $\Phi(\sigma_2)$ and $\Psi(\alpha_1)$ vs $\Psi(\alpha_2)$, where $\sigma_1 \in \{0.05, 0.1, 0.15\}$, $\sigma_2 \in \Sigma$, $\alpha_1 \in \{0.02, 0.03, \alpha_{max}\}$ and $\alpha_2 \in \mathcal{A}$. In Experiment 3 in the main text, realizations of these models were observed in $W_1 = [0,1]^2$. Now we consider $W_2 = [0,2]^2$, which means that the number of points observed in a pattern is on average four times higher. The size and composition of training and testing data are the same as in Experiment 3 in the main text.

Dissimilarity measures and classification scenarios We choose $\mathrm{d}_{int}[g]$ and $\mathrm{d}_{sup}[g]$ to be the dissimilarity measures. For $\sigma_1 \in \{0.05, 0.1, 0.15\}$ and $\sigma_2 \in \Sigma$, we denote

$$\begin{aligned}\varphi_{W_2,int}[\sigma_1,\sigma_2](\cdot) &= \varphi_{NW}(\,\cdot \mid \mathcal{T}(\sigma_1,\sigma_2,20,20,W_2), \mathrm{d}_{int}[g], \mathcal{K}, h_{k_{LCV}}),\\ \varphi_{W_2,sup}[\sigma_1,\sigma_2](\cdot) &= \varphi_{NW}(\,\cdot \mid \mathcal{T}(\sigma_1,\sigma_2,20,20,W_2), \mathrm{d}_{sup}[g], \mathcal{K}, h_{k_{LCV}}).\end{aligned}$$

Note that the notation suppresses the dependence on the functional summary characteristic. The choice of K and $h_{k_{LCV}}$ is the same as in Experiment 1 in the main text, and in $\mathcal{T}(\sigma_1,\sigma_2,20,20,W_2)$ we emphasize that the training data (and also the testing data) consist of point patterns observed in W_2. The average misclassification rates corresponding to these scenarios will be compared with those for $\varphi_{W_1,int}[\sigma_1,\sigma_2]$ and $\varphi_{W_1,sup}[\sigma_1,\sigma_2]$ from Experiment 3 in the main text. For $\alpha_1 \in \{0.02, 0.03, \alpha_{max}\}$ and $\alpha_2 \in \mathcal{A}$, the corresponding scenarios $\varphi_{W_2,int}[\alpha_1,\alpha_2]$ and $\varphi_{W_2,sup}[\alpha_1,\alpha_2]$ are considered.

Results for the Thomas process We expect to observe an improved performance of the proposed classifiers, compared to Experiment 3 in the main text, since they are based on empirical pair correlation functions observed from a higher number of points. Indeed, Fig. S8.16 shows that, in terms of the average misclassification rate, $\varphi_{W_2,int}$ outperforms $\varphi_{W_1,int}$ for all values of σ_2 and all choices of σ_1. For $\sigma_1 = 0.05$ almost perfect classification is achieved, except for $\sigma_2 = 0.04$ or 0.06, the two neighbouring values of σ_1. The variability of the individual misclassification rates corresponding to the I repetitions of the experiment is extremely low, resulting in a very narrow pointwise envelope in this case. For $\sigma_1 = 0.1$ and 0.15, the performance is significantly improved compared to using W_1 in Experiment 3 in the main text, most notably for values of σ_2 not very close to σ_1. Furthermore, the pointwise envelopes are more narrow here than those reported in Fig. 10 in the main text.

The results of this experiment for $\varphi_{W_2,sup}$ are given in Fig. S8.17. The difference between the performance of $\varphi_{W_2,int}$ and $\varphi_{W_2,sup}$ is rather negligible; see Fig. S8.18. For numerical values of the average misclassification rate, see Tab. S8.9 and Tab. S8.10.

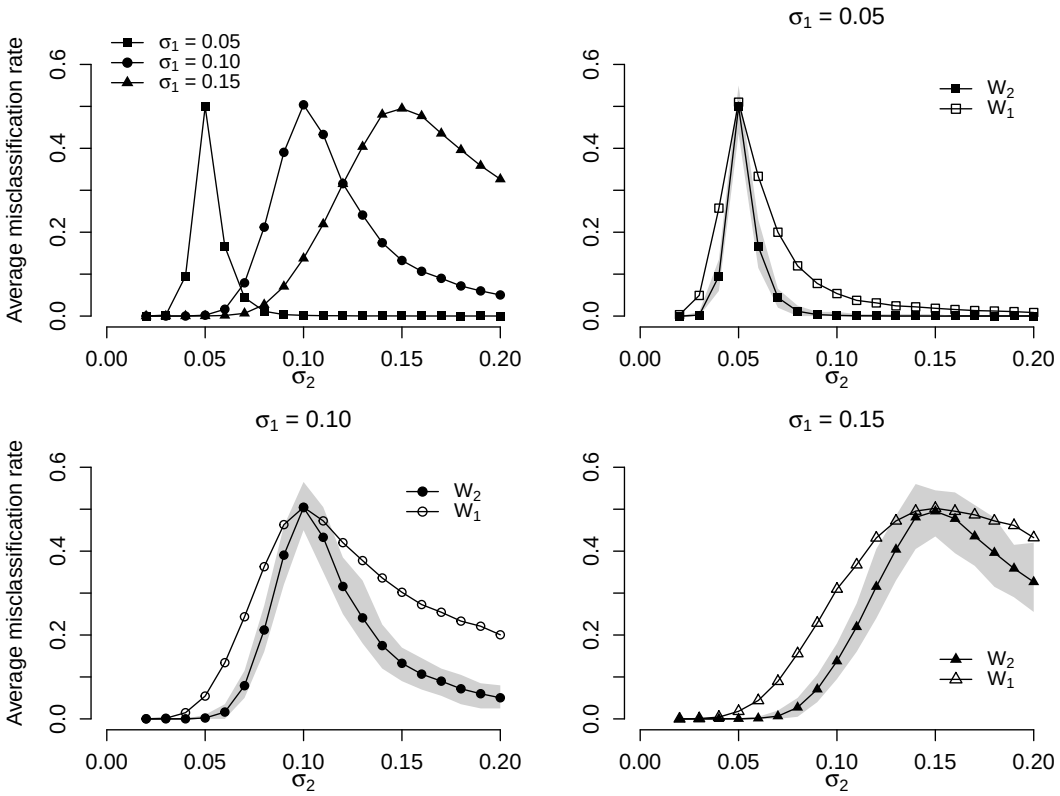

Figure S8.16: For $\sigma_1 \in \{0.05, 0.1, 0.15\}$, average misclassification rates $\bar{\gamma}\left(\varphi_{W_2,int}[\sigma_1, \sigma_2]\right)$ are plotted as functions of σ_2. These values (including the 90% pointwise envelopes) are compared with $\bar{\gamma}\left(\varphi_{W_1,int}[\sigma_1, \sigma_2]\right)$ from Experiment 3 in the main text (empy symbols).

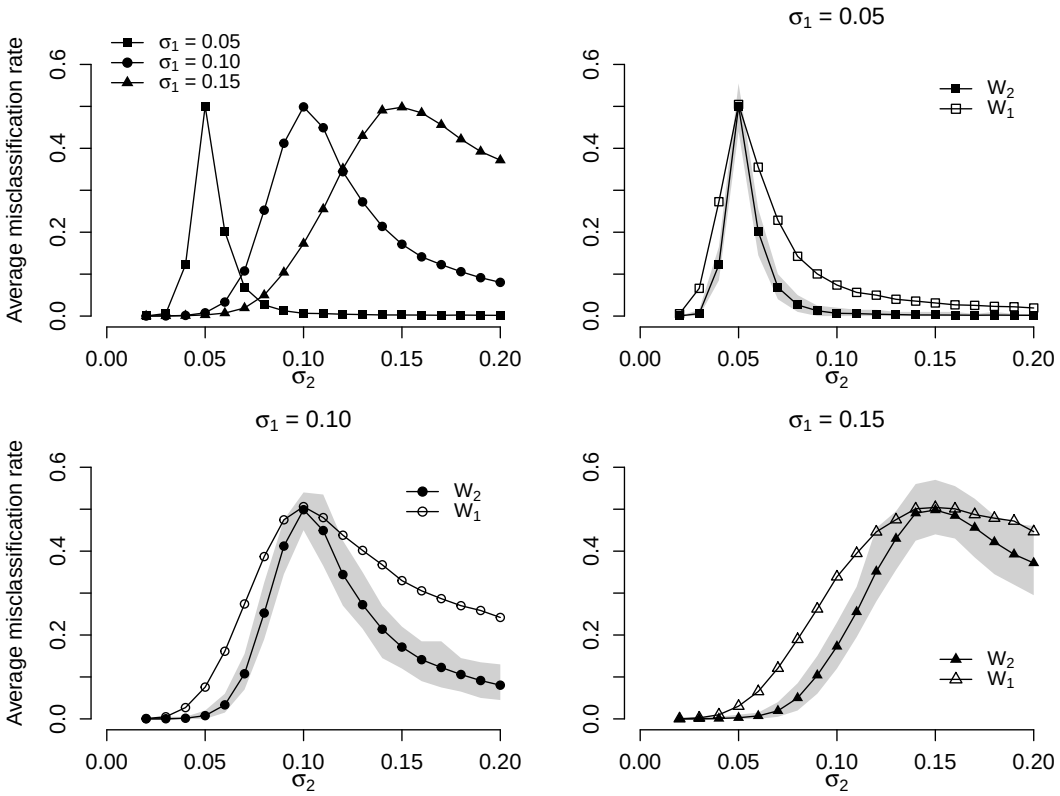

Figure S8.17: For $\sigma_1 \in \{0.05, 0.1, 0.15\}$, average misclassification rates $\bar{\gamma}\left(\varphi_{W_2,sup}[\sigma_1, \sigma_2]\right)$ are plotted as functions of σ_2. These values, including the 90% pointwise envelopes, are compared with $\bar{\gamma}\left(\varphi_{W_1,sup}[\sigma_1, \sigma_2]\right)$ from Experiment 3 in the main text (empy symbols).

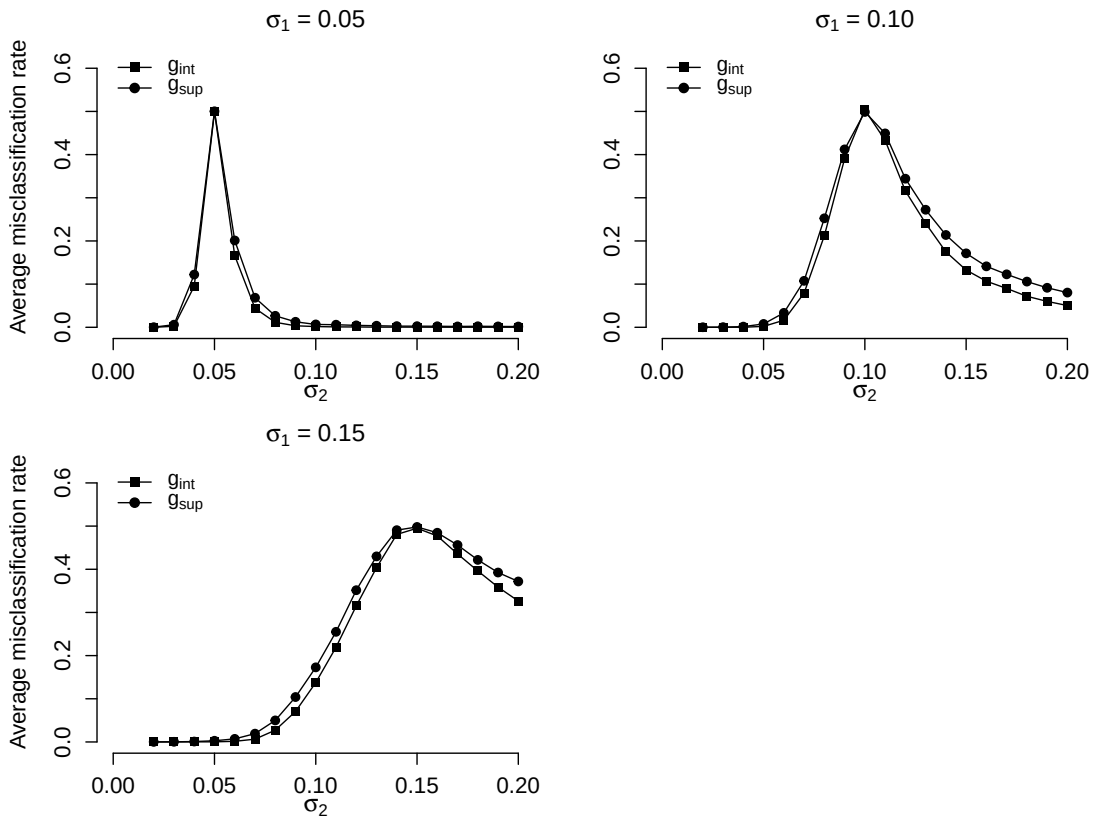

Figure S8.18: Binary classification $\Phi(\sigma_1)$ vs $\Phi(\sigma_2)$, $\sigma_1 \in \{0.05, 0.10, 0.15\}$, $\sigma_2 \in \Sigma$. This figure is a counterpart of Fig. S7.15; all of the realizations are now observed in the observation window $W = [0, 2]^2$.

	Average misclassification rate					
$\sigma_2 \setminus \sigma_1$	0.05		0.10		0.15	
0.02	0.003	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
0.03	0.049	0.002	0.002	< 0.001	0.001	< 0.001
0.04	0.257	0.095	0.015	< 0.001	0.004	< 0.001
0.05	0.510	0.500	0.054	0.002	0.018	< 0.001
0.06	0.333	0.166	0.134	0.016	0.044	0.001
0.07	0.200	0.044	0.244	0.079	0.090	0.007
0.08	0.120	0.012	0.363	0.212	0.156	0.027
0.09	0.078	0.004	0.463	0.390	0.229	0.071
0.10	0.054	0.002	0.505	0.504	0.310	0.138
0.11	0.037	0.001	0.472	0.433	0.368	0.219
0.12	0.031	0.001	0.420	0.316	0.432	0.315
0.13	0.024	< 0.001	0.378	0.241	0.472	0.404
0.14	0.022	< 0.001	0.336	0.175	0.496	0.481
0.15	0.019	< 0.001	0.302	0.133	0.502	0.495
0.16	0.016	< 0.001	0.273	0.107	0.495	0.477
0.17	0.013	< 0.001	0.254	0.090	0.487	0.435
0.18	0.012	< 0.001	0.233	0.072	0.472	0.396
0.19	0.011	< 0.001	0.221	0.060	0.461	0.359
0.20	0.008	< 0.001	0.201	0.050	0.432	0.326

Table S8.9: Binary classification $\Phi(\sigma_1)$ vs $\Phi(\sigma_2)$, $\sigma_1 \in \{0.05, 0.10, 0.15\}$, $\sigma_2 \in \Sigma$. Values of the average misclassification rate calculated from the realizations in the unit square (left) are compared with the average misclassification rate calculated from the realizations in the observation window $[0, 2]^2$ (right). The underlying dissimilarity measure is $\mathrm{d}_{int}[g]$.

	Average misclassification rate					
$\sigma_2 \setminus \sigma_1$	0.05		0.10		0.15	
0.02	0.006	< 0.001	0.001	< 0.001	< 0.001	< 0.001
0.03	0.066	0.006	0.005	< 0.001	0.002	< 0.001
0.04	0.273	0.122	0.027	0.002	0.010	0.001
0.05	0.505	0.500	0.076	0.008	0.030	0.003
0.06	0.355	0.201	0.161	0.033	0.065	0.007
0.07	0.229	0.068	0.274	0.108	0.121	0.019
0.08	0.143	0.027	0.387	0.252	0.190	0.050
0.09	0.101	0.013	0.475	0.412	0.262	0.104
0.10	0.074	0.007	0.506	0.499	0.339	0.173
0.11	0.056	0.006	0.480	0.449	0.395	0.255
0.12	0.050	0.004	0.438	0.344	0.446	0.352
0.13	0.040	0.004	0.402	0.272	0.475	0.430
0.14	0.036	0.003	0.367	0.214	0.501	0.491
0.15	0.031	0.003	0.330	0.171	0.504	0.498
0.16	0.027	0.002	0.305	0.141	0.501	0.485
0.17	0.026	0.002	0.288	0.123	0.487	0.456
0.18	0.023	0.002	0.270	0.106	0.479	0.422
0.19	0.022	0.002	0.259	0.092	0.472	0.392
0.20	0.019	0.002	0.242	0.080	0.446	0.372

Table S8.10: Binary classification $\Phi(\sigma_1)$ vs $\Phi(\sigma_2)$, $\sigma_1 \in \{0.05, 0.10, 0.15\}$, $\sigma_2 \in \Sigma$. This table is a counterpart of Tab. S8.9 for $\mathrm{d}_{sup}[g]$.

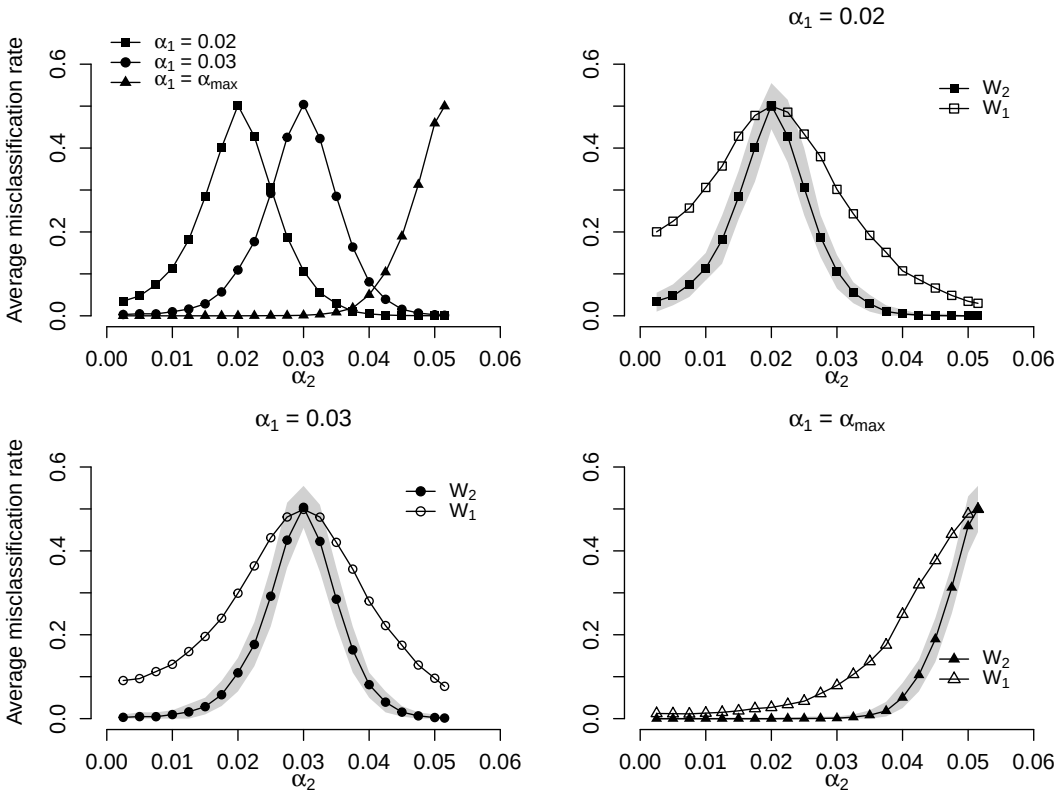

Figure S8.19: For $\alpha_1 \in \{0.02, 0.03, \alpha_{max}\}$, average misclassification rates $\bar{\gamma}\left(\varphi_{W_2,int}[\alpha_1, \alpha_2]\right)$ are plotted as functions of α_2. These values (including the 90% pointwise envelopes) are compared with $\bar{\gamma}\left(\varphi_{W_1,int}[\alpha_1, \alpha_2]\right)$ from Experiment 3 in the main text (empty symbols).

Results for the Gaussian determinantal process For the classification $\Psi(\alpha_1)$ vs $\Psi(\alpha_2)$, we can draw similar conclusions as in the previous paragraph. Fig. S8.19 shows that, disregarding the value of α_1, improved performance of $\varphi_{W_2,int}$ was achieved compared to that based on W_1, again most notably for values of α_2 not very close to α_1. The difference between the performance of $\varphi_{W_2,int}$ and $\varphi_{W_2,sup}$ is negligible. For numerical values of the average misclassification rate, see Tab. S8.11 and Tab. S8.12.

Summary The increasing number of points in the observation window increases the accuracy of the estimator $\widehat{g}$ of the pair correlation function. This has a positive impact on the performance of φ, for both situations $\Phi(\sigma_1)$ vs $\Phi(\sigma_2)$ and $\Psi(\alpha_1)$ vs $\Psi(\alpha_2)$. As observed in Experiment 3 in the main text, the satisfactory classification between models with strong interactions and weak interactions can be obtained using patterns with on average 120 points. With the number of points increasing to 480 on average, the average misclassification rate is reduced in most situations, except for the extreme cases where the values of model parameters are very close to each other.

	Mean					
$\alpha_2 \setminus \alpha_1$	0.0200		0.0300		α_{max}	
0.0025	0.200	0.035	0.091	0.003	0.012	0
0.0050	0.225	0.047	0.096	0.005	0.012	0
0.0075	0.257	0.074	0.112	0.005	0.011	0
0.0100	0.306	0.114	0.130	0.010	0.013	0
0.0125	0.357	0.181	0.160	0.016	0.015	0
0.0150	0.428	0.284	0.196	0.029	0.019	< 0.001
0.0175	0.478	0.401	0.239	0.057	0.024	< 0.001
0.0200	0.500	0.502	0.299	0.109	0.027	< 0.001
0.0225	0.486	0.427	0.364	0.177	0.034	< 0.001
0.0250	0.433	0.305	0.431	0.292	0.042	< 0.001
0.0275	0.380	0.188	0.481	0.426	0.060	0.001
0.0300	0.301	0.107	0.499	0.504	0.079	0.001
0.0325	0.243	0.055	0.481	0.423	0.105	0.003
0.0350	0.192	0.029	0.420	0.285	0.136	0.008
0.0375	0.151	0.010	0.357	0.164	0.176	0.019
0.0400	0.107	0.005	0.280	0.081	0.249	0.050
0.0425	0.087	0.002	0.222	0.039	0.319	0.104
0.0450	0.066	0.001	0.175	0.016	0.377	0.190
0.0475	0.049	< 0.001	0.128	0.007	0.440	0.313
0.0500	0.035	< 0.001	0.096	0.003	0.488	0.459
α_{max}	0.030	< 0.001	0.077	0.001	0.499	0.500

Table S8.11: Binary classification $\Psi(\alpha_1)$ vs $\Psi(\alpha_2)$, $\alpha_1 \in \{0.02, 0.03, \alpha_{max}\}$, $\alpha \in \mathcal{A}$. Values of the average misclassification rate calculated from the realizations in the unit square (left) are compared with the average misclassification rate computed from the realizations in the observation window $[0, 2]^2$ (right). The underlying dissimilarity measure is $\mathrm{d}_{int}[g]$.

	Average misclassification rate					
$\alpha_2 \setminus \alpha_1$	0.0200		0.0300		α_{max}	
0.0025	0.197	0.035	0.087	0.003	0.013	0
0.0050	0.222	0.044	0.095	0.005	0.013	< 0.001
0.0075	0.254	0.070	0.110	0.004	0.011	< 0.001
0.0100	0.307	0.114	0.131	0.009	0.014	< 0.001
0.0125	0.358	0.181	0.155	0.015	0.016	< 0.001
0.0150	0.427	0.281	0.197	0.028	0.020	< 0.001
0.0175	0.471	0.402	0.241	0.057	0.025	< 0.001
0.0200	0.500	0.501	0.302	0.109	0.027	< 0.001
0.0225	0.483	0.428	0.358	0.177	0.035	< 0.001
0.0250	0.434	0.306	0.430	0.290	0.043	< 0.001
0.0275	0.374	0.189	0.475	0.425	0.063	0.001
0.0300	0.300	0.107	0.500	0.494	0.081	0.002
0.0325	0.244	0.057	0.482	0.419	0.106	0.004
0.0350	0.189	0.028	0.420	0.285	0.142	0.009
0.0375	0.152	0.010	0.361	0.167	0.179	0.020
0.0400	0.109	0.005	0.283	0.081	0.257	0.053
0.0425	0.086	0.002	0.225	0.040	0.328	0.108
0.0450	0.067	0.001	0.181	0.017	0.385	0.196
0.0475	0.048	0.001	0.130	0.007	0.443	0.320
0.0500	0.035	< 0.001	0.101	0.003	0.488	0.462
α_{max}	0.030	< 0.001	0.080	0.002	0.503	0.498

Table S8.12: Binary classification $\Psi(\alpha_1)$ vs $\Psi(\alpha_2)$, $\alpha_1 \in \{0.02, 0.03, \alpha_{max}\}$, $\alpha_2 \in \mathcal{A}$. This table is a counterpart of Tab. S8.11 for $\mathrm{d}_{sup}[g]$.

S9 Real-data example

This section provides additional visual material to the real-data example in the main text. Ternary classification is considered; a collection of 68 point patterns from [8] representing the centres of intramembranous particles located in the mitochondrial membranes of the HeLa cell line is examined.

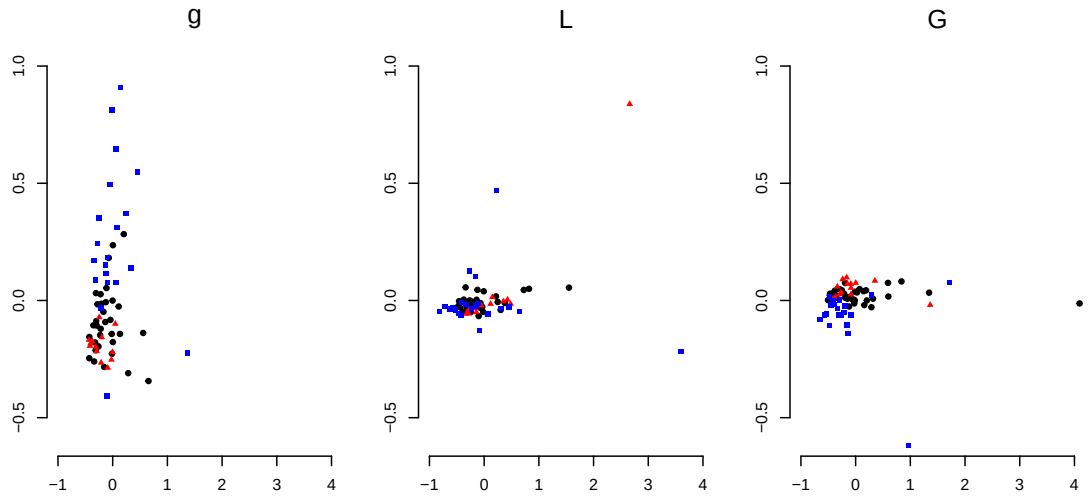

Figure S9.20: Ternary classification of a real dataset containing 68 point patterns representing the centres of intramembranous particles located in mitochondrial membranes of the HeLa cell line. The cell line has been observed in three different environments to study the mitochondrial metabolism: normal conditions (33 patterns), after exposition to sodium acid (14 patterns) and after exposition to rotenone (21 patterns). For detailed information, see [8]. The plots correspond to the visualization of dissimilarities in the dataset for $\mathrm{d}_{int}[g]$ (left), $\mathrm{d}_{int}[L]$ (middle) and $\mathrm{d}_{int}[G]$ (right). Patterns from the first group (normal conditions) are visualised by black circles, patterns from the second group (sodium acid) are visualised by red triangles and patterns from the third group (rotenone) are represented by blue squares. The two-dimensional coordinates of the points of the plot are determined by the multidimensional scaling so that the distance of a pair of points in the plot is approximately proportional to the dissimilarity between the underlying pair of point patterns.

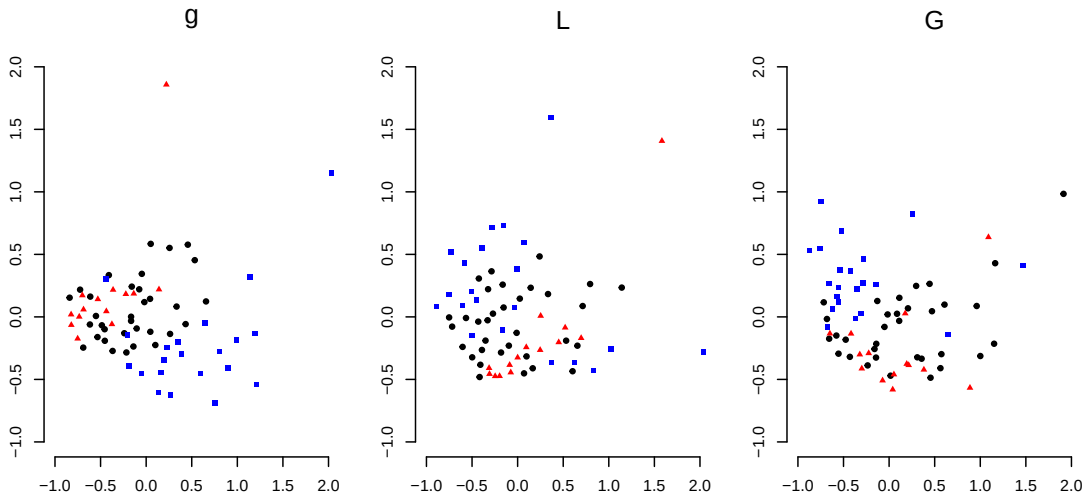

Figure S9.21: Ternary classification of a real dataset containing 68 point patterns representing the centres of intramembranous particles located in mitochondrial membranes of the HeLa cell line. The cell line has been observed in three different environments to study the mitochondrial metabolism: normal conditions (33 patterns), after exposition to sodium acid (14 patterns) and after exposition to rotenone (21 patterns). For detailed information, see [8]. Same situation as in Fig. S9.20, the plots now corresponds to $\mathrm{d}_{sup}[g]$ (left), $\mathrm{d}_{sup}[L]$ (middle) and $\mathrm{d}_{sup}[G]$ (right).

S10 Selected asymptotic results

Here, we give an overview of selected asymptotic results for the kernel regression classifier, available in the literature. In equation (5) in the main text, the estimate of the conditional probabilities $\left\{p_{\mathrm{g}}(\cdot),\ \mathrm{g} \in \overline{\mathrm{G}}\right\}$ is based on the collection of the labeled point patterns $\mathcal{T}_N$. Suppose now that we have a collection of random pairs $\mathbb{X} = \{(X_i, Y_i),\ i = 1, 2, \ldots, N\}$, that are independent and identically distributed as the random pair (X, Y). We are interested in the consistency of the estimator $\widehat{p}_{\mathrm{g}}(\cdot \mid \mathbb{X})$, when N grows to infinity. We abbreviate the almost complete convergence as *a.co.* The rate of convergence is denoted by $\mathcal{O}$. Accordingly, the rate of the almost complete convergence is denoted by $\mathcal{O}_{a.co.}$.

We focus on the results presented in [3, Sect. 8.5]. Thus, we follow the notation established therein. We are in the context of functional data analysis; the random pair (X, Y) is so that X takes values in the semi-metric space (E, δ), E is a set of functions, and Y is the label variable with values in $\overline{\mathrm{G}}$. Let C_E^0 be the classes of continuous functions in E, i.e., functions $s : E \longrightarrow \mathbb{R}$ such that

$$\lim_{\delta(\mathcal{X},\mathcal{X}')\to 0} s\left(\mathcal{X}'\right) = s\left(\mathcal{X}\right).$$

Similarly, let Lip_E^{β} denote the class of Lipschitz continuous functions on E, that is functions $s : E \longrightarrow \mathbb{R}$ such that

$$\exists C > 0\ \forall \mathcal{X}, \mathcal{X}' \in E\ \left|s\left(\mathcal{X}\right) - s\left(\mathcal{X}'\right)\right| < C\delta\left(\mathcal{X}, \mathcal{X}'\right).$$

For $\mathcal{X}$, an element of E and for $\epsilon > 0$, let us denote by $\varphi_{\mathcal{X}}(\varepsilon)$ the probability that X is included in an open ball centered at $\mathcal{X}$, with radius ϵ. In what follows, we may need to add further conditions on these probabilities

$$\exists C > 0\ \exists \varepsilon_0 > 0\ \forall \varepsilon < \varepsilon_0\ \ \int_0^{\varepsilon} \varphi_{\mathcal{X}}(\eta)\ \mathrm{d}\eta > 0. \tag{1}$$

We say that the kernel function $\mathrm{K} : \mathbb{R} \longrightarrow [0, \infty)$ is of *type I* if there exist $\exists\ 0 < C_1 < C_2 < \infty$ such that for all $u \in \mathbb{R}$

$$C_1 \mathbf{1}\left\{0 \leq u \leq 1\right\} \leq \mathrm{K}(u) \leq C_2 \mathbf{1}\left\{0 \leq u \leq 1\right\}.$$

K is of *type II* if its support equals $[0, 1]$, the first derivation K' exists on $[0, 1]$ and

$$\exists\ -\infty < C_2 < C_1 < 0\ \forall u \in \mathbb{R}\ \ C_2 \leq \mathrm{K}'(u) \leq C_1.$$

Theorem S10.1 (Theorem 8.1 in [3])**.** *Fix $\mathcal{X} \in E$. Let $\{h_n,\ n \in \mathbb{N}\}$ be a sequence of non-negative numbers such that*

$$\lim_{n\to\infty} h_n = 0 \text{ and } \lim_{n\to\infty} \frac{logn}{n\varphi_{\mathcal{X}}(h_n)} = 0.$$

Let the probabilities $\left\{p_{\mathrm{g}}(\mathcal{X}),\ \mathrm{g} \in \overline{\mathrm{G}}\right\}$ belong to C_E^0. Let $\varphi_{\mathcal{X}}(\varepsilon) > 0$, for all $\varepsilon > 0$. Moreover, let one of these two conditions be fulfilled: K *is a type I kernel or* K *is a type II kernel and* (1) *holds. Then, for* $\mathrm{g} \in \overline{\mathrm{G}}$

$$\widehat{p}_{\mathrm{g}}\left(\mathcal{X} \mid \mathbb{X}, \delta, \mathrm{K}, h_n\right) \underset{n\to\infty}{\longrightarrow} p_{\mathrm{g}}\left(\mathcal{X}\right) \quad a.co.$$

Theorem S10.2 (Theorem 8.2 in [3]). *Fix $\mathcal{X} \in E$. Let the probabilities $\left\{p_{\mathrm{g}}(\mathcal{X}),\ \mathrm{g} \in \overline{\mathrm{G}}\right\}$ belong to Lip_E^{β}. Then, under the assumptions of Theorem S10.1, we find that the difference $\widehat{p}_{\mathrm{g}}\left(\mathcal{X} \mid \mathbb{X}, \delta, \mathrm{K}, h_n\right) - p_{\mathrm{g}}\left(\mathcal{X}\right)$ is of the order of*

$$\mathcal{O}\left(h_n^{\beta}\right) + \mathcal{O}_{a.co}\left(\sqrt{\frac{\log n}{n\varphi_{\mathcal{X}}(\varepsilon)}}\right).$$

We have several remarks accompanying these two theorems. First, note that the assumption of continuity or Lipschitz continuity applies to the conditional probabilities $\{p_{\mathrm{g}}(\mathcal{X}), \mathrm{g} \in \overline{\mathrm{G}}\}$ not to the elements of the space (E, δ). Thus, the fact that some of the mentioned estimated point process characteristics have jumps is not a drawback in the use of these two theorems. Second, the semi-metric δ is always nonnegative. Thus, these theorems work with the so-called asymmetrical version of the standard kernel functions with support $[0, 1]$ and appropriate normalization. Then, the asymmetric uniform kernel is of type I and the asymmetrical Epanechnikov kernel is of type II. Third, both theorems can be immediately used with $E \subset \mathcal{N}_{lf}$, where $\mathcal{N}_{lf}$ is the space of all locally finite subsets of $\mathbb{R}^2$, and δ a semi-metric built above one of the functional characteristics for point processes. However, verification of some of the assumptions can be a challenging task even for simple point process models.

The results concerning the mean-square consistency of the Nadaraya-Watson type estimators of a regression function with a real response variable Y and the explanatory variable X in a general metric space (E, δ) are given in [4]. Note that above, we are within the scope of the Nadaraya-Watson type estimators of a regression function with a categorical response variable Y and the explanatory variable X in a functional semi-metric space (E, δ).

References

[1] Baddeley A, Gill R (1997) Kaplan-Meier estimators of distance distributions for spatial point processes, Annals of Statistics, 25, 263-292

[2] Baddeley A, Rubak E, Turner R (2015) Spatial Point Patterns: Methodology and Applications with R., Chapman and Hall/CRC Press, Boca Raton

[3] Ferraty F, Vieu P (2006) Nonparametric functional data analysis. Theory and practice. Springer-Verlag, New York

[4] Forzani L, Fraiman R, Llop P (2012) Consistent nonparametric regression for functional data under the Stone–Besicovitch conditions, IEEE Trans Inf Theory, 58, 6697-6708

[5] Illian J, Penttinen A, Stoyan H, Stoyan D (2004) Statistical Analysis and Modelling of Spatial Point Patterns. Wiley, Chichester

[6] Møller J, Waagepetersen RP (2004) Statistical Inference and Simulation for Spatial Point Processes. Chapman & Hall/CRC, Boca Raton

[7] R Core Team (2017) R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/

[8] Schladitz K, Särkkä A, Pavenstädt I, Haferkamp O, Mattfeldt T (2003) Statistical analysis of intramembranous particles using freeze fracture specimens, Journal of Microscopy, 211, 137-153