

Raport

Michał Milewski

5/16/2020

Wstęp

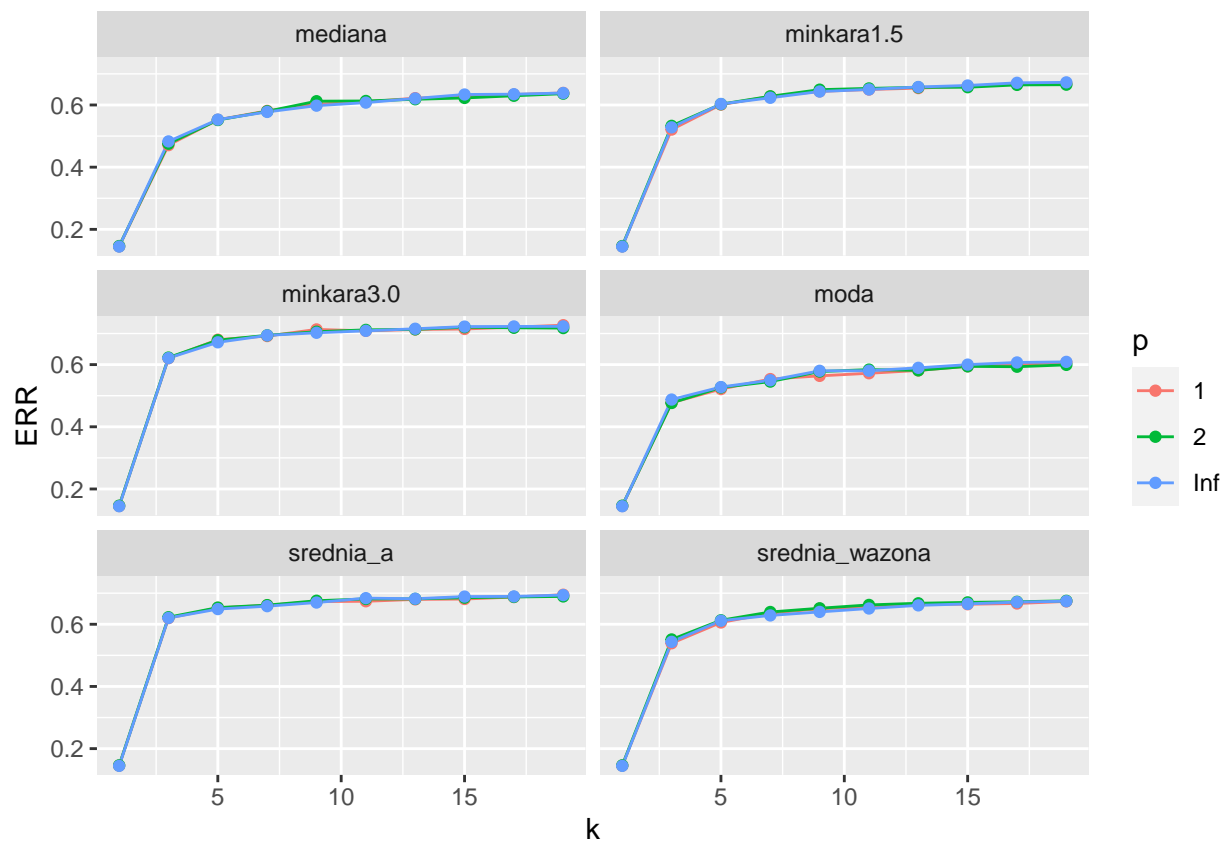
Bierzący raport dotyczy porównania działania funkcji knn z użyciem różnych funkcji agregujących etykiety oraz funkcji dostępnych jako gotowe metody. Testowane metody to **randomForest::randomForest** , **MASS::polr** i **e1071::svm**.

Zbiór danych *abalone*

Poniżej znajdują się wykresy obrazujące następujące dane:

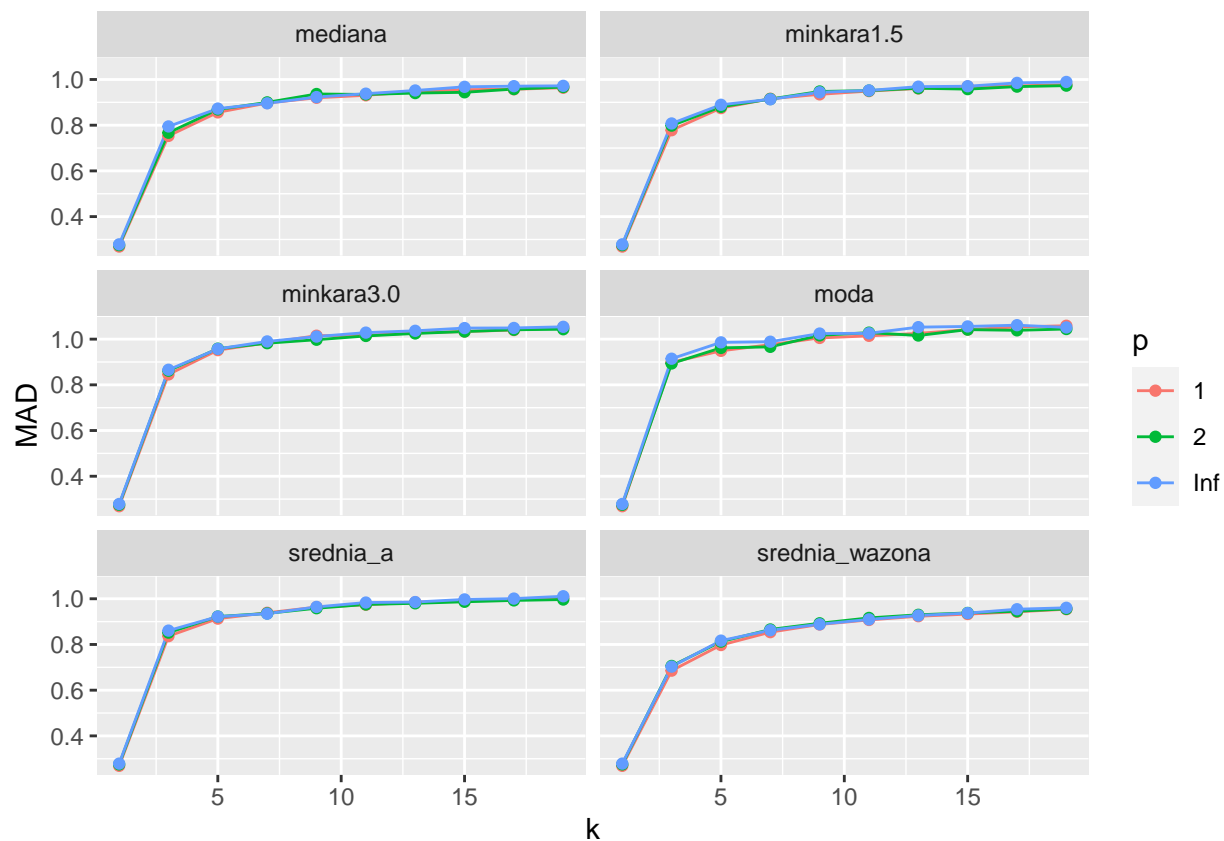
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że błąd ERR zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



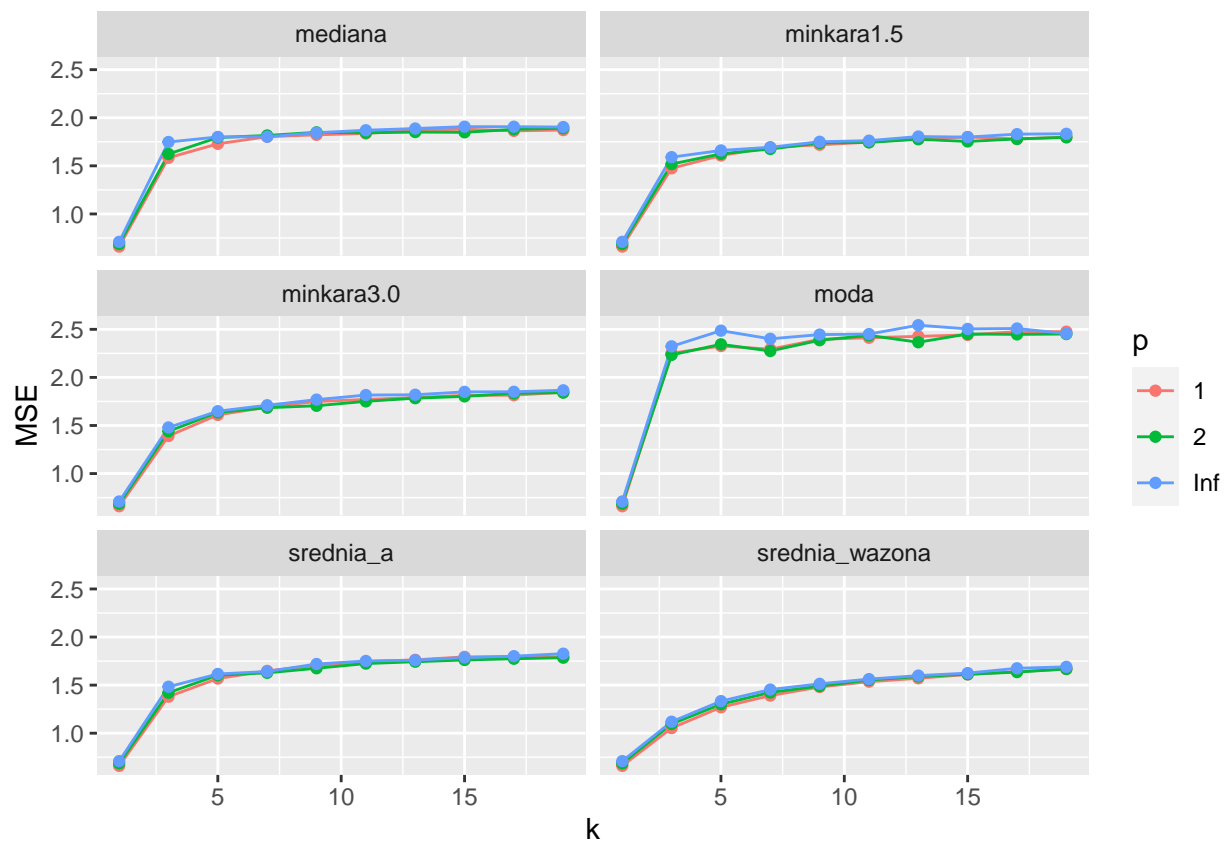
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem bezwzględnych w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem średniokwadratowym w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. W tym przypadku widać że minimalnie można zaobserwować niższe błędy dla metryki L1. W tym przypadku *moda* cechuje się największym błędem średniokwadratowym.



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru danych można zaobserwować że najlepiej radzi sobie funkcja *randomForest* z najmniejszymi wszystkimi błędami. Błąd ERR tej funkcji jest porównywalny z błędem 1-nn. Błędy pozostałych funkcji są porównywalne z błędami knn dla $k > 10$ oraz użyciem funkcji agregującej *moda*.

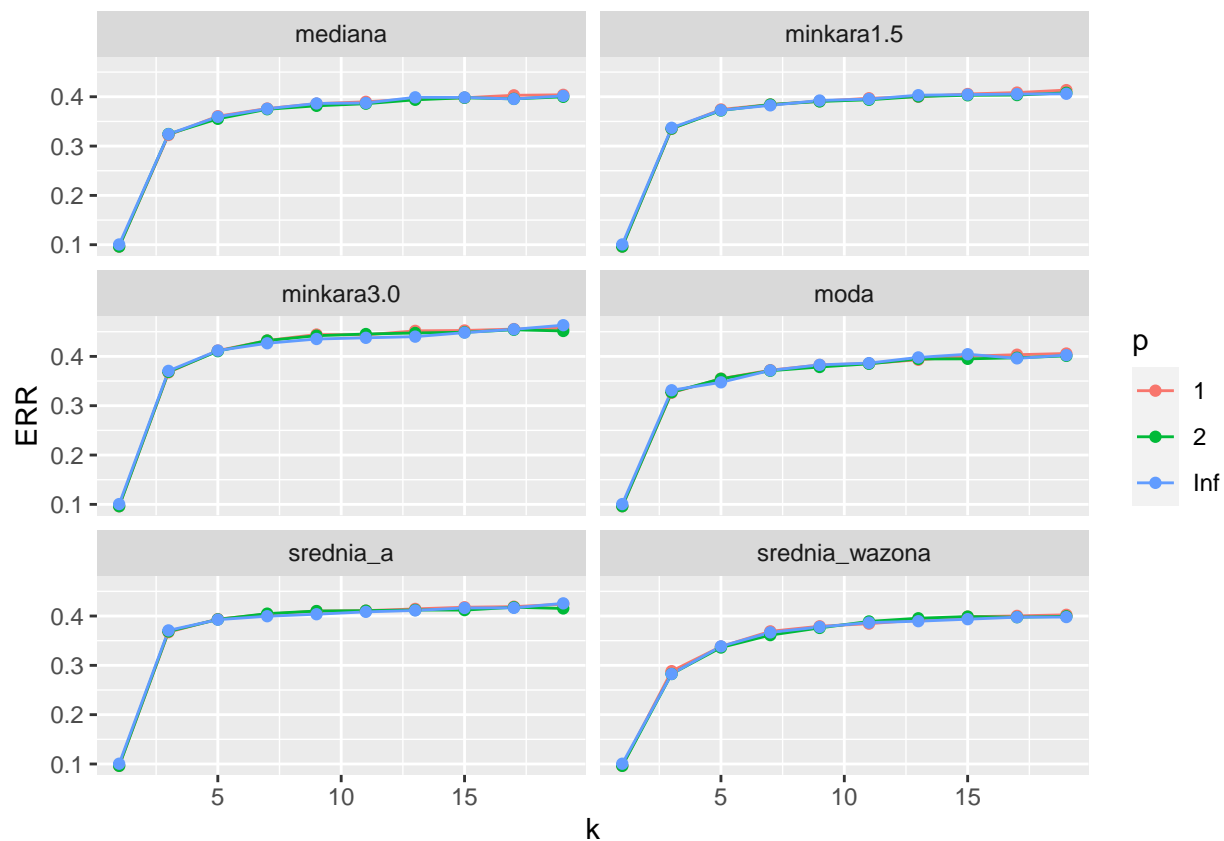
name	ERR	MAD	MSE
randomForest	0.1369361	0.2310223	0.5173501
polr	0.6667504	1.1364671	2.5331722
svm	0.6282059	1.1098814	2.6932691

Zbiór danych *abalone_ord*

Poniżej znajdują się wykresy obrazujące następujące dane:

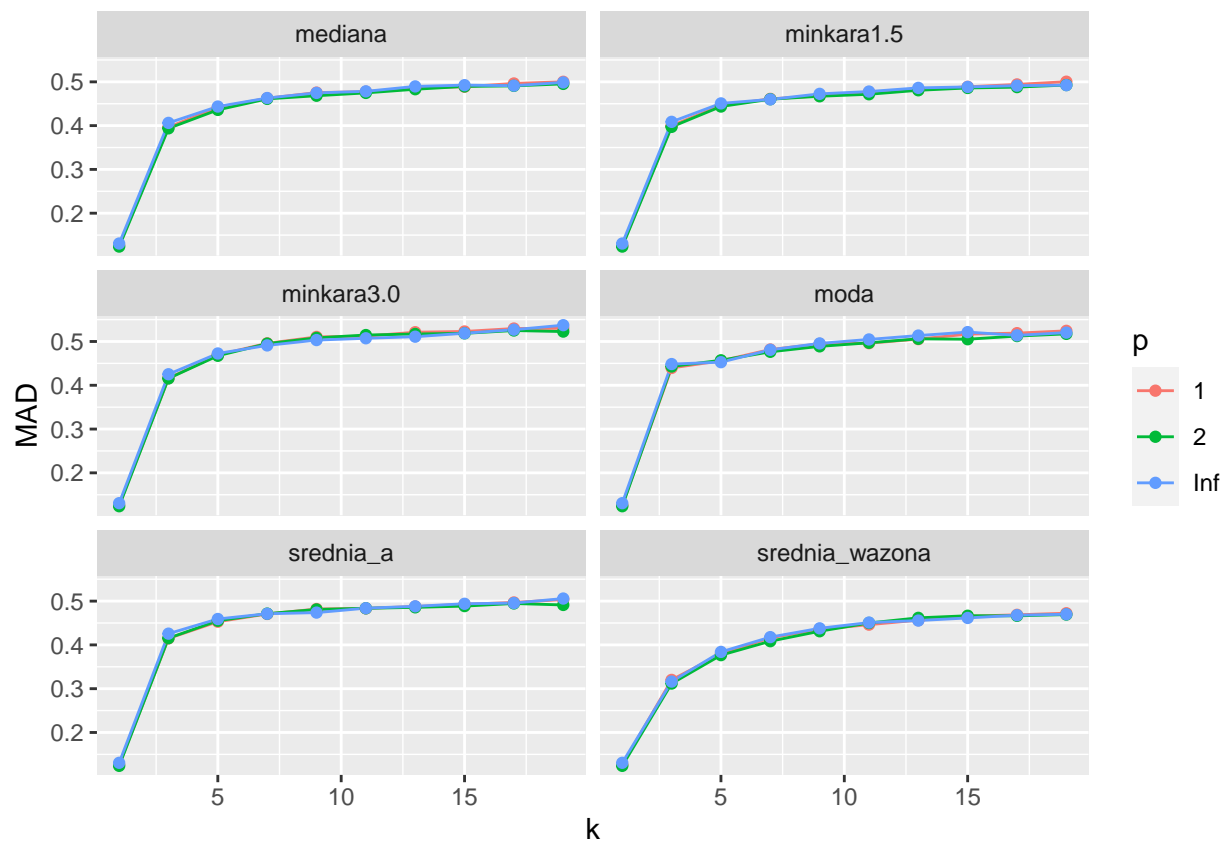
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że błąd ERR zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



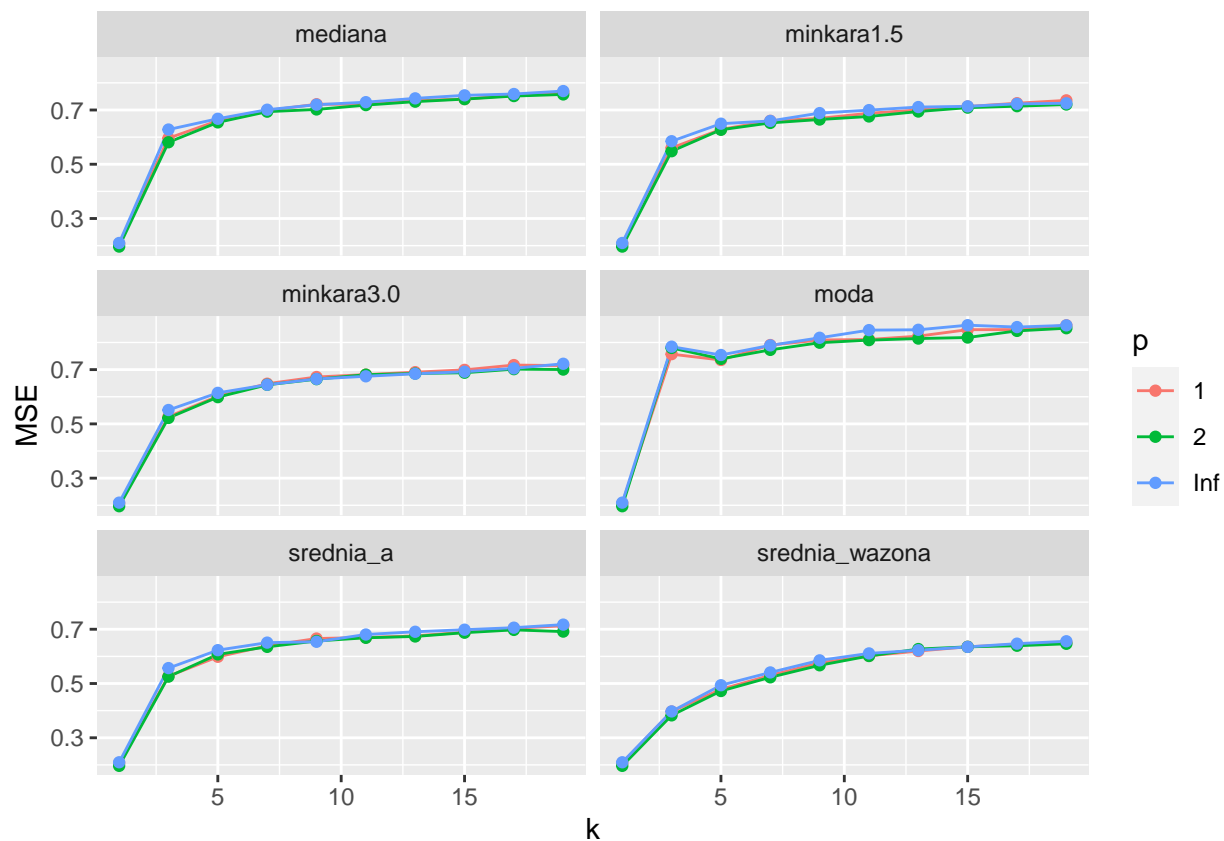
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem bezwzględnych w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem średniokwadratowym w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. W tym przypadku *moda* cechuje się największym błędem średniokwadratowym.



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru danych można zaobserwować że najlepiej radzi sobie funkcja *randomForest* z najmniejszymi wszystkimi błędami. Błąd ERR tej funkcji jest porównywalny z błędem 1-nn. Błędy pozostałych funkcji są porównywalne z błędami knn dla $k > 10$ oraz użyciem funkcji agregującej *moda*.

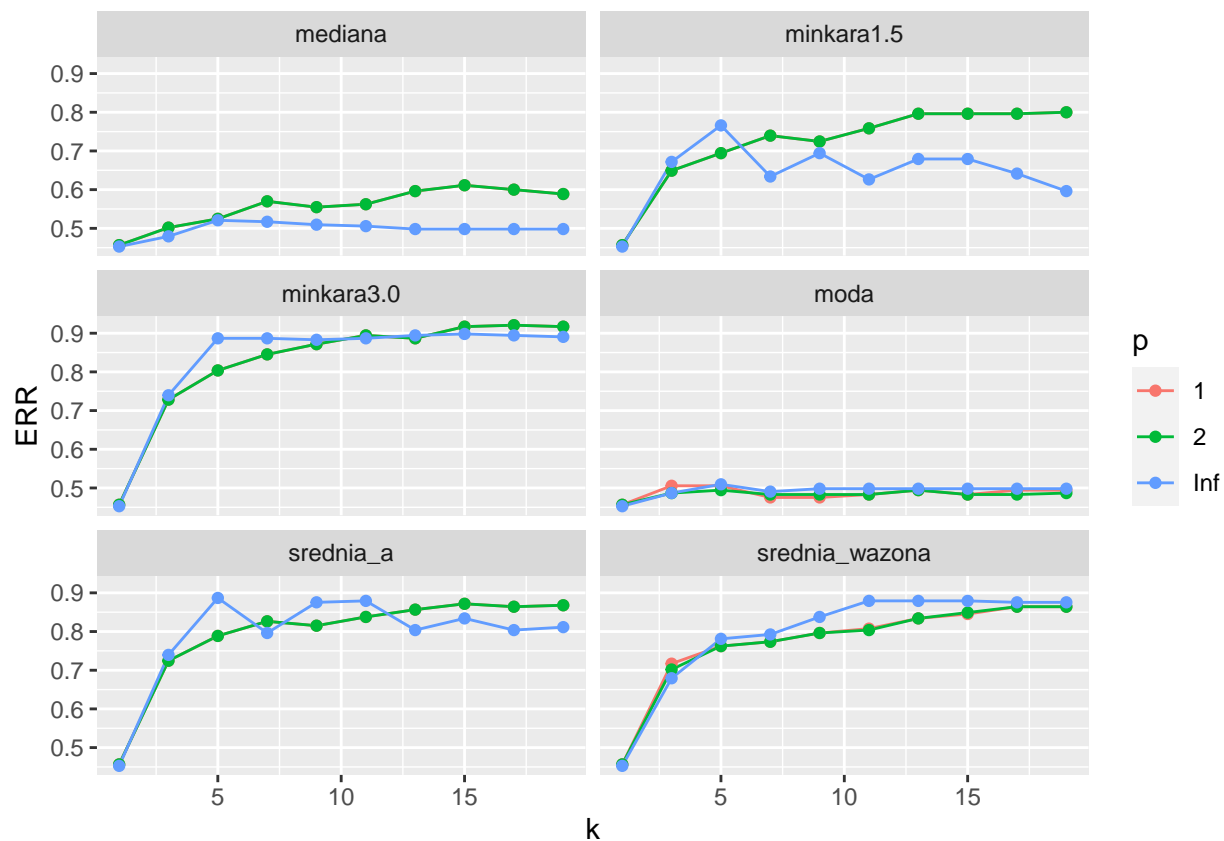
name	ERR	MAD	MSE
randomForest	0.08905796	0.1108438	0.1697338
polr	0.44409793	0.5472807	0.8360003
svm	0.42086955	0.5527697	0.9487081

Zbiór danych *affairs*

Poniżej znajdują się wykresy obrazujące następujące dane:

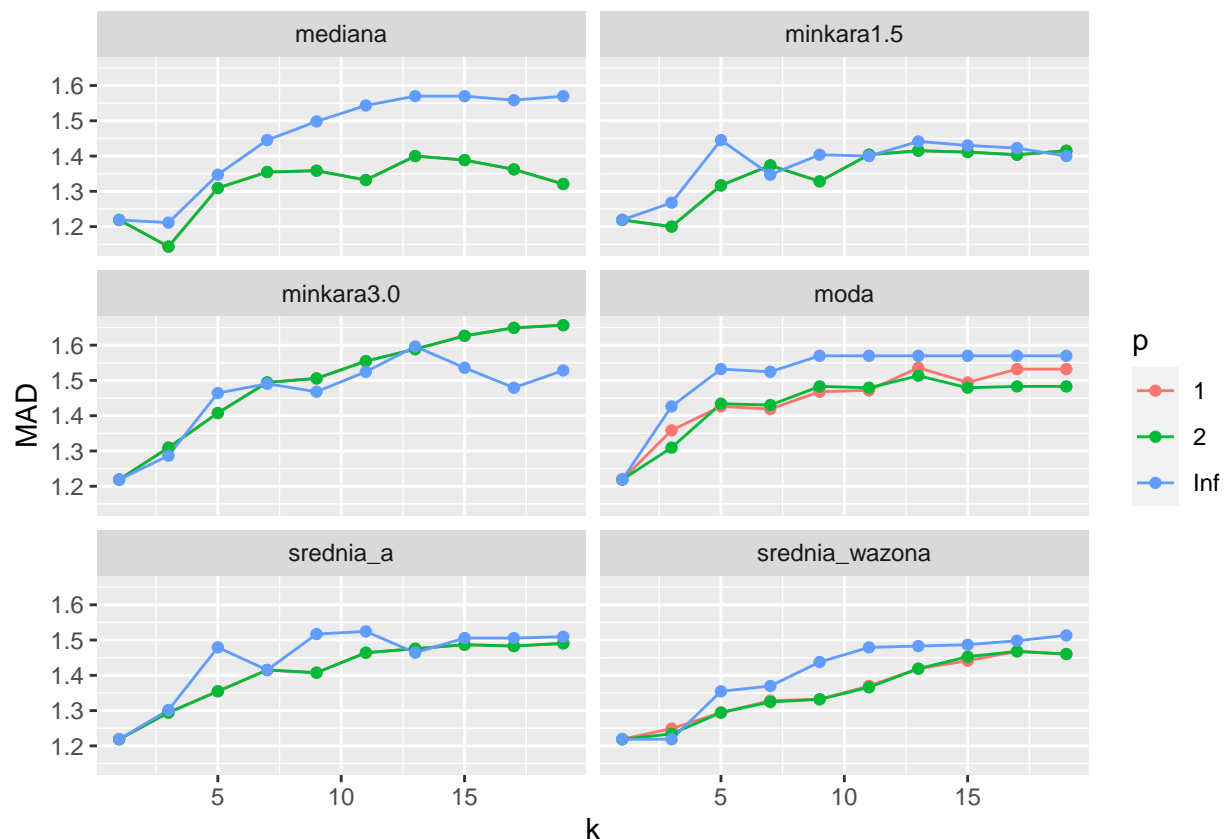
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Wyjątkiem od tej reguły jest tu użycie funkcji agregującej minkara 1.5, która to przy użyciu metryki L_∞ , zaczyna maleć dla $k > 3$. Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. Najlepsze wyniki oferuje funkcja *moda*.



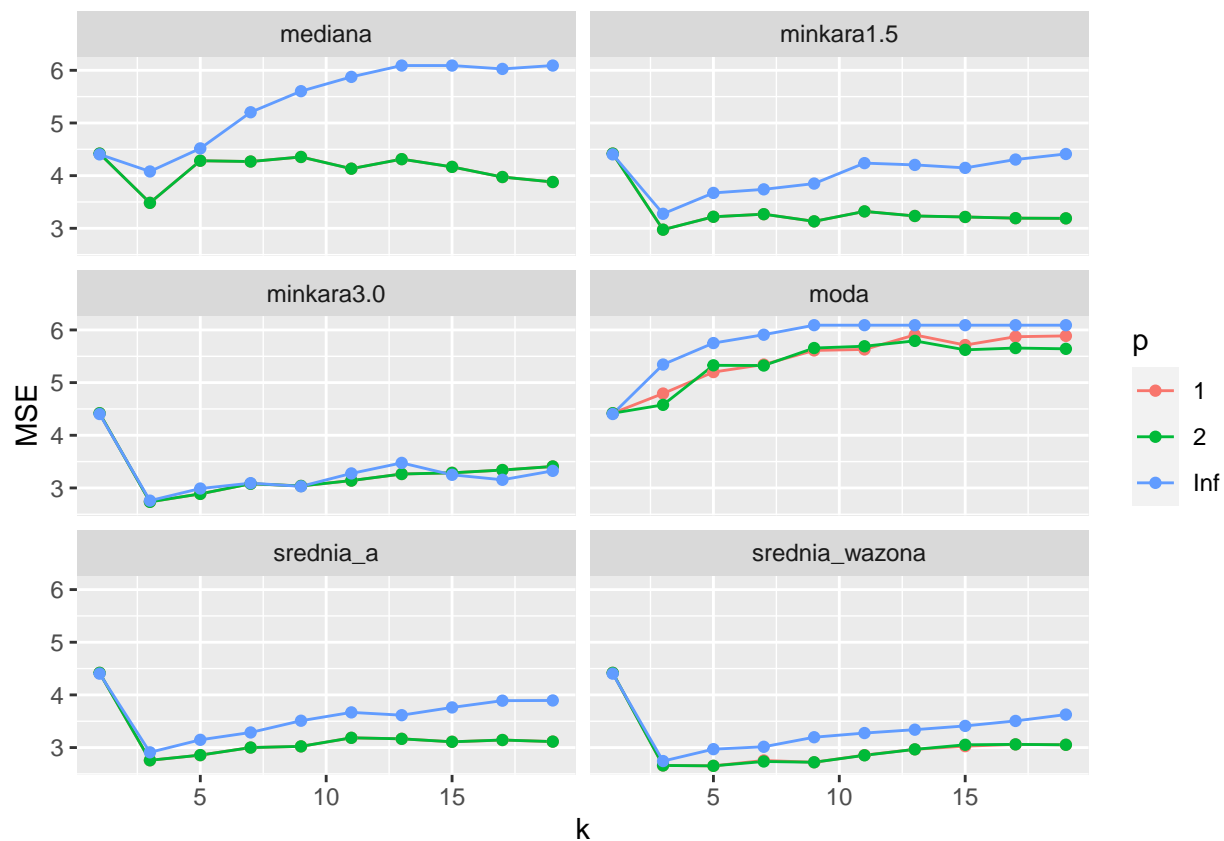
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem bezwzględnych w tym przypadku cechuje się funkcja *minkara1.5*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3, choć w przypadku funkcji *srednia_wazona* jest on zdecydowanie łagodniejszy.



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem średniokwadratowym w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k , tak jest dla funkcji agregującej *moda*. Jednak dla innych funkcji błąd dla $k = 1$ jest duży, a najmniejszy dla $k=3$ i później rośnie. Najgwałtowniejszy spadek można zaobserwować dla k od 1 do 3. W tym przypadku *moda* cechuje się największym błędem średniokwadratowym. tu również widzimy oewne rozbieżności między błędami w przypadku użycia różnych metryk. Metryka L_∞ daje największe błędy w tym przypadku.



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru błędy wszystkich użytych funkcji są porównywalne z nieznacznie mniejszymi przy funkcji *randomForest*

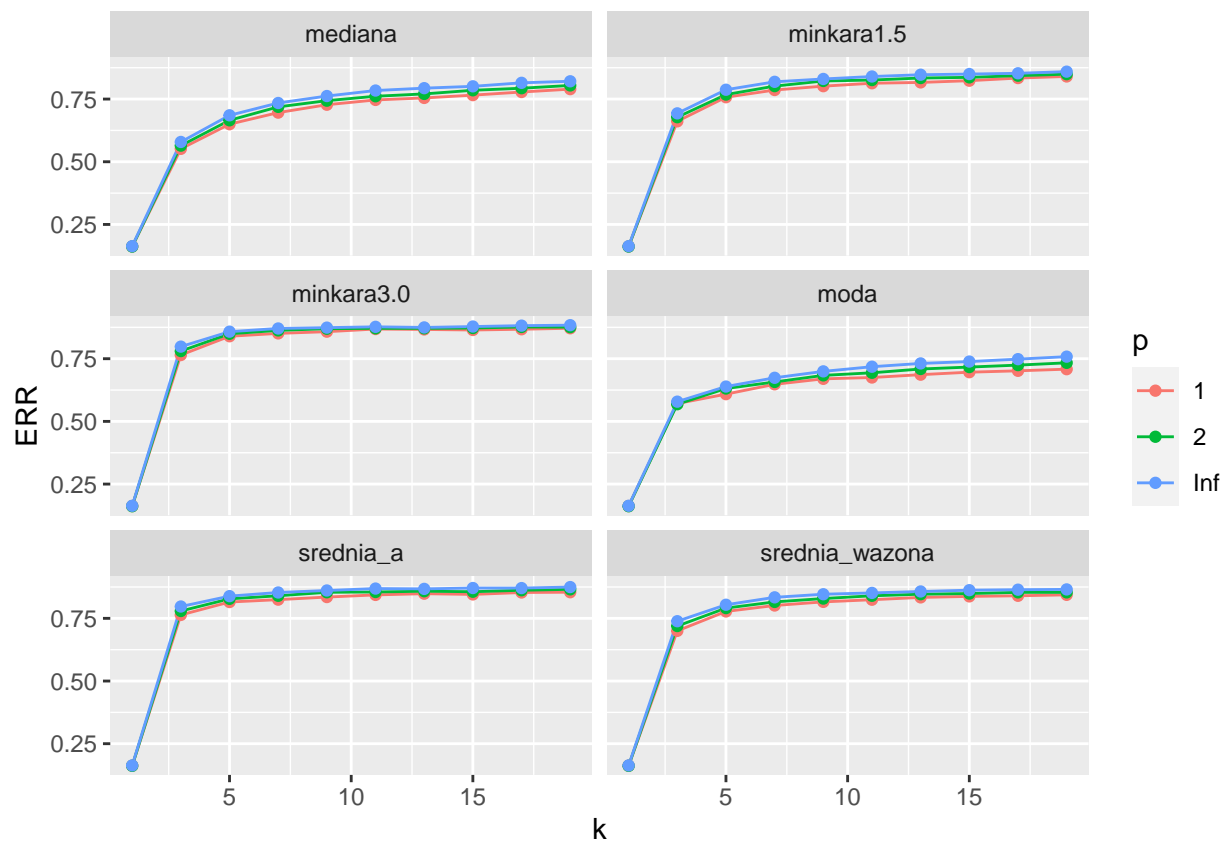
name	ERR	MAD	MSE
randomForest	0.4339623	1.264151	4.683019
polr	0.4981132	1.483019	5.611321
svm	0.4943396	1.547170	6.000000

Zbiór danych *aileron*s

Poniżej znajdują się wykresy obrazujące następujące dane:

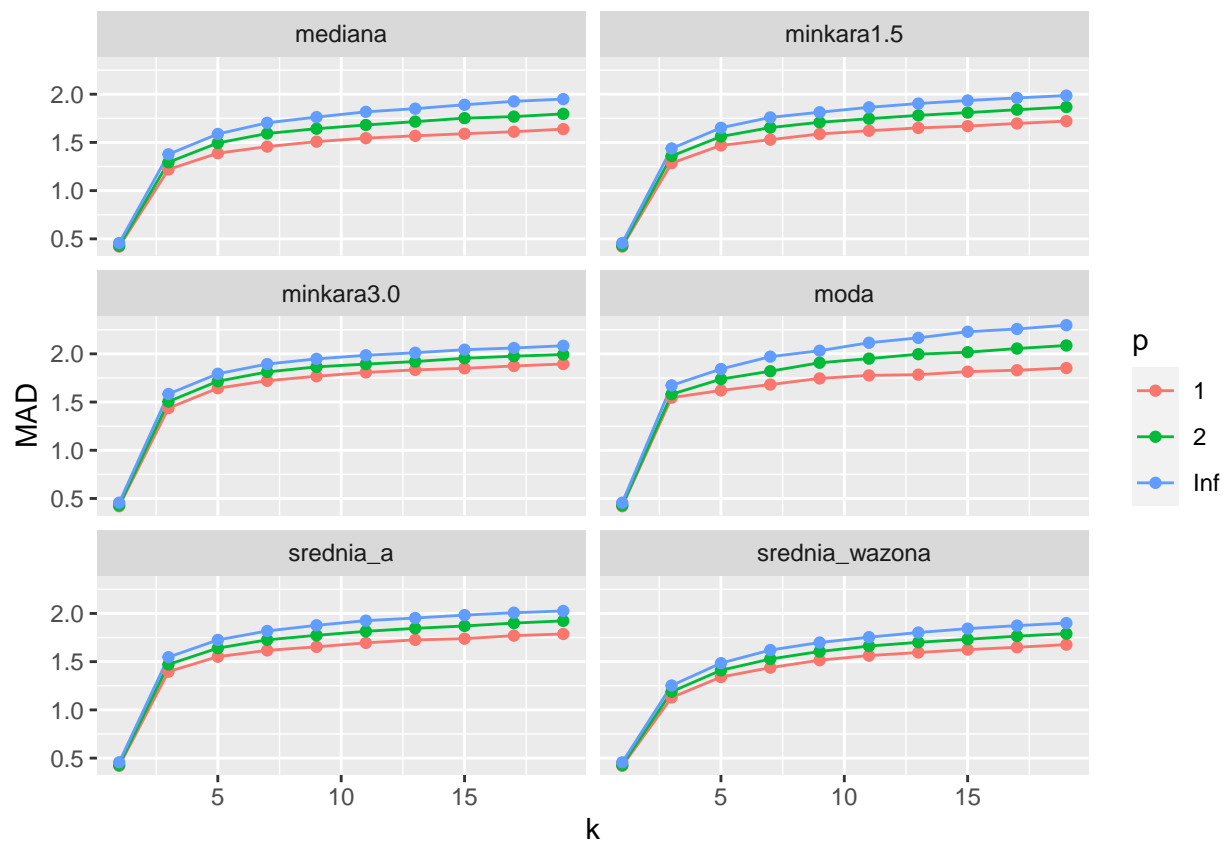
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że błąd ERR zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



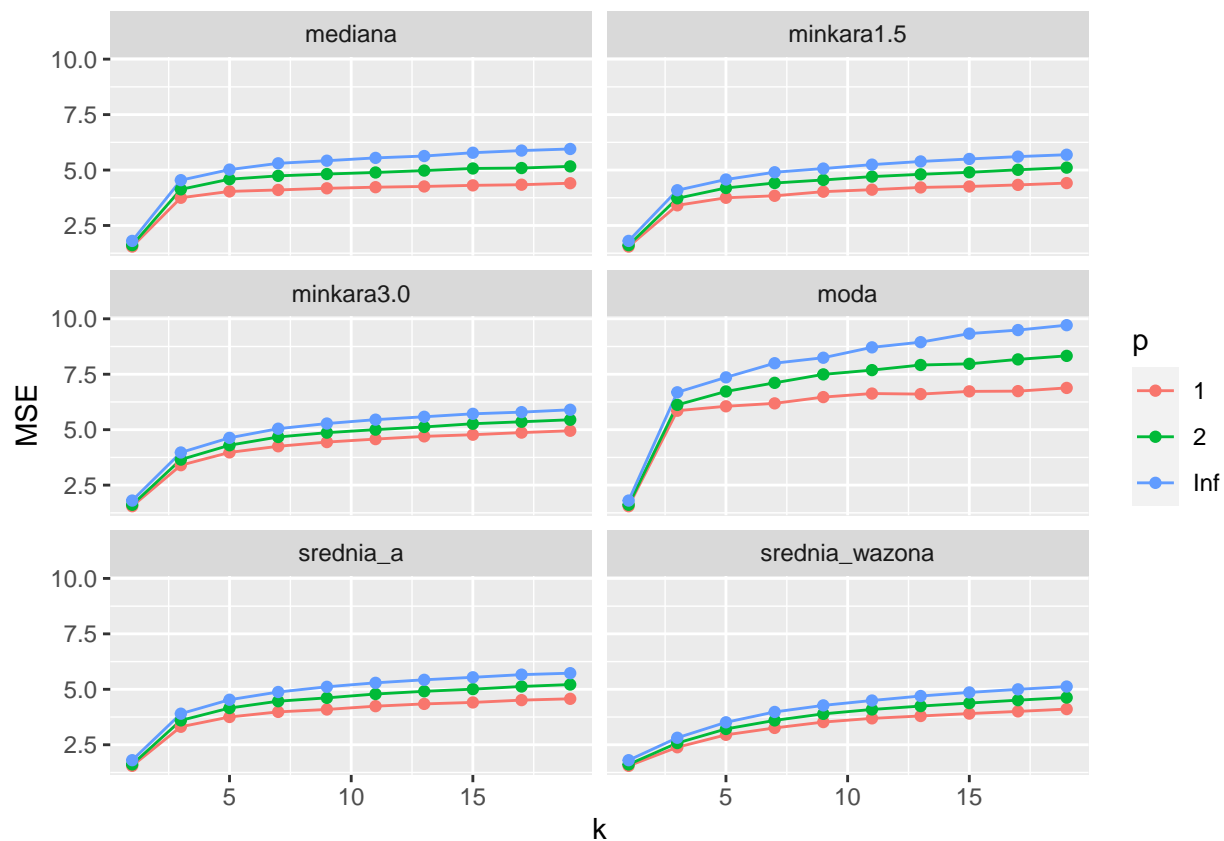
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem bezwzględnych w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . W tym przypadku widać że można zaobserwować niższe błędy dla metryki L1. Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem średniokwadratowym w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. W tym przypadku widać że można zaobserwować niższe błędy dla metryki L1, trochę większe dla L2 i największe dla L_∞ . W tym przypadku moda cechuje się największym błędem średniokwadratowym.



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru danych można zaobserwować że najlepiej radzi sobie funkcja *randomForest* z najmniejszymi wszystkimi błędami. Błąd ERR tej funkcji jest porównywalny z błędem 1-nn. Błędy ERR pozostałych funkcji są porównywalne z błędami knn dla $k > 10$ oraz użyciem funkcji agregującej *moda*. Wszystkie te funkcje cechują zauważalnie mniejszymi błędami MAD i MSE.

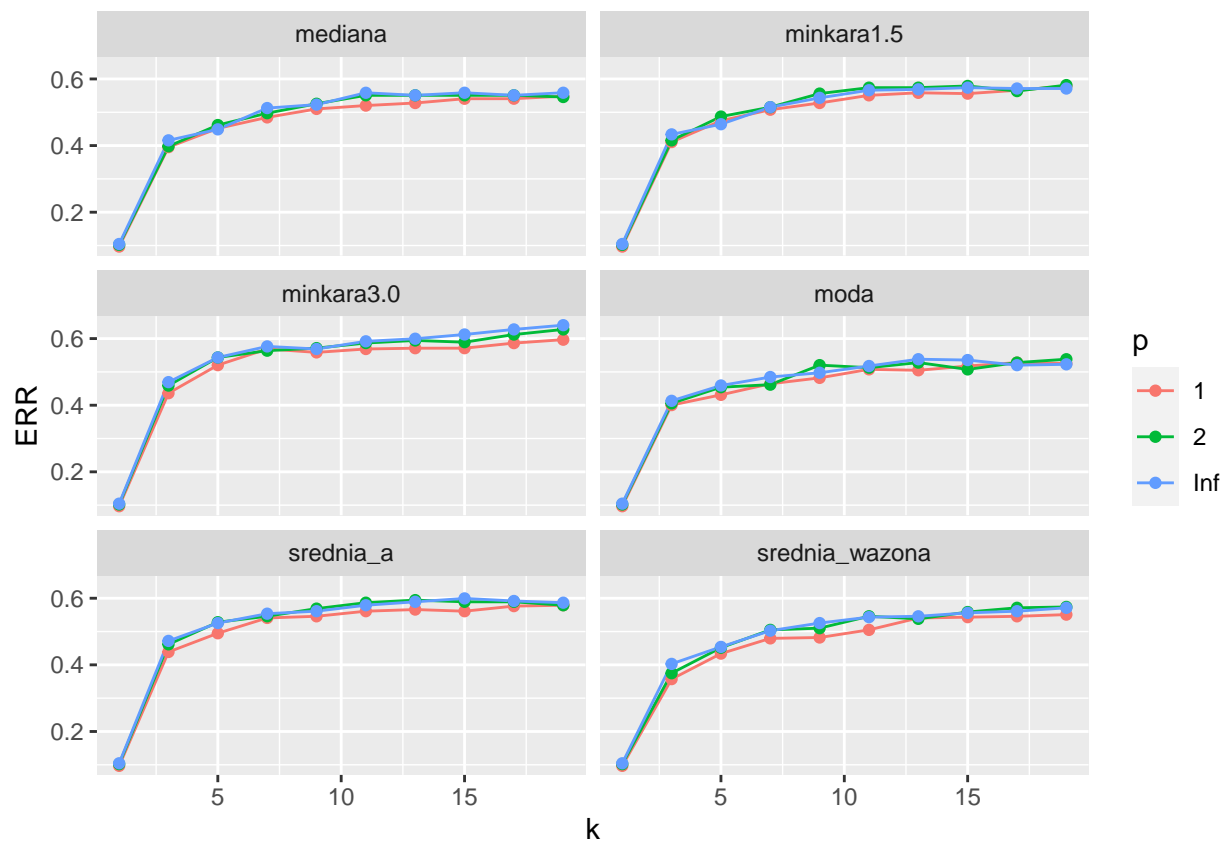
name	ERR	MAD	MSE
randomForest	0.1172768	0.1877281	0.4015999
polr	0.5641600	0.8153501	1.4994497
svm	0.4470312	0.7423899	1.7314726

Zbiór danych *auto_ord*

Poniżej znajdują się wykresy obrazujące następujące dane:

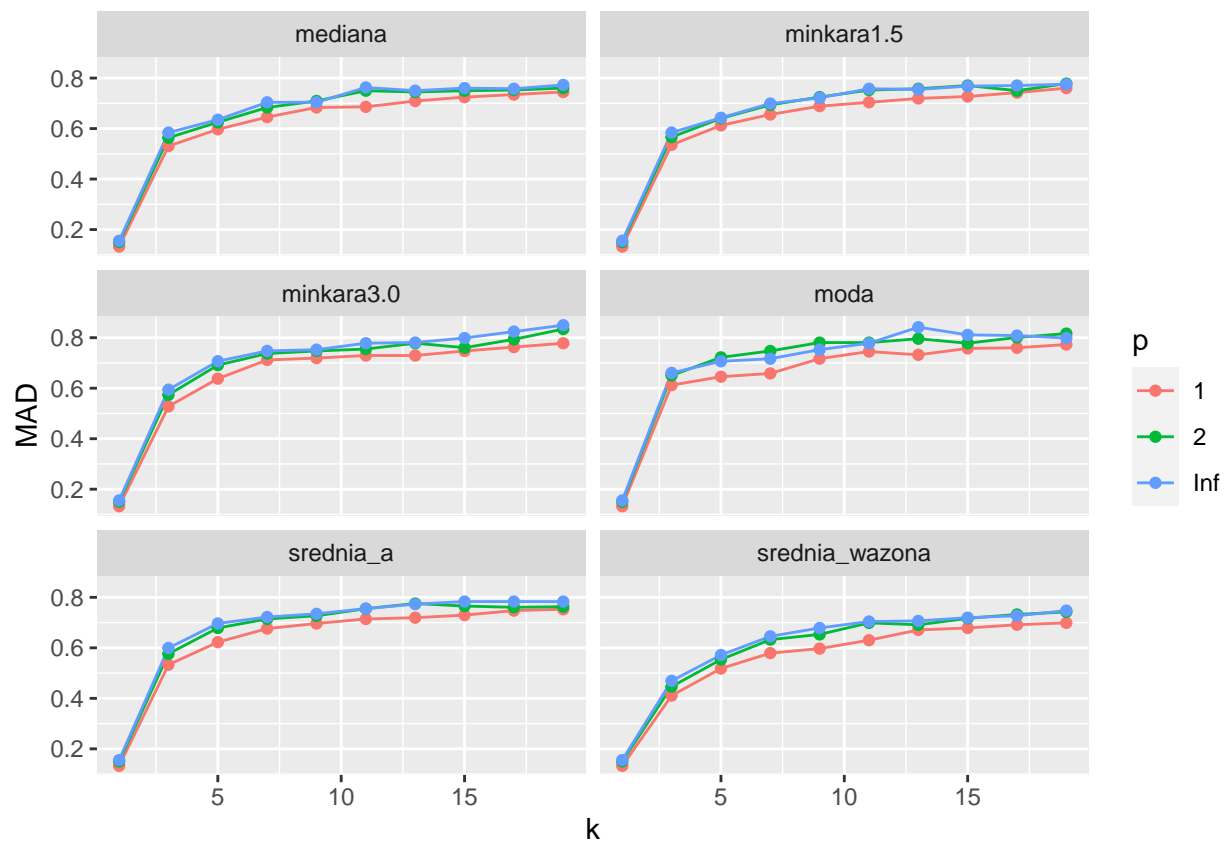
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że błąd ERR zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. Widać też że dla niektórych k zauważalnie zmniejsza się błąd dla metryki L1



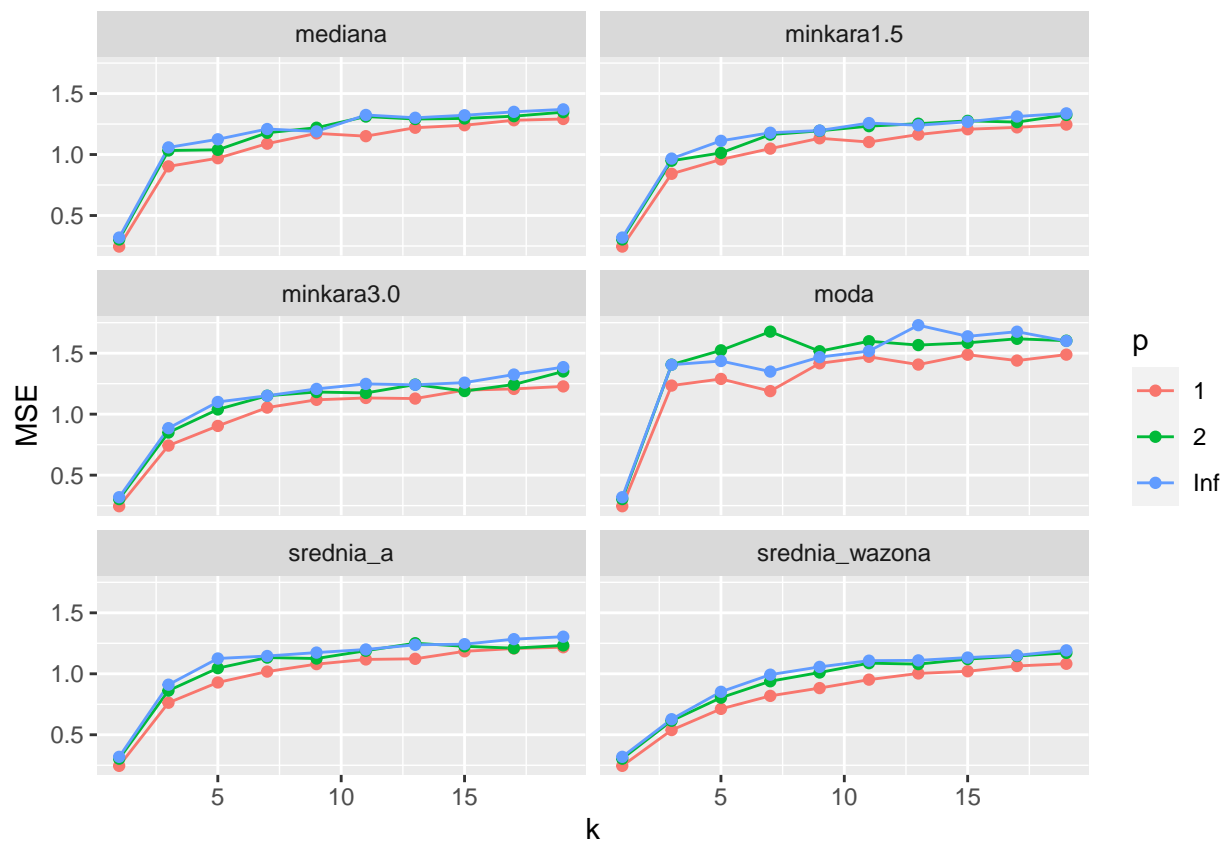
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem bezwzględnych w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . W tym przypadku widać że można zaobserwować niższe błędy dla metryki L1. Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem średniokwadratowym w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. W tym przypadku widać że można zaobserwować niższe błędy dla metryki L1, trochę większe dla L2 i największe dla L_∞ . W tym przypadku *moda* cechuje się największym błędem średniokwadratowym.



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru danych można zaobserwować że najlepiej radzi sobie funkcja *randomForest* z najmniejszymi wszystkimi błędami. Błąd ERR tej funkcji jest porównywalny z błędem 1-nn. Błędy ERR pozostałych funkcji są porównywalne z błędami knn dla $k > 3$ oraz użyciem funkcji agregującej *moda*. Wszystkie te funkcje cechują zauważalnie mniejszymi błędami MAD i MSE.

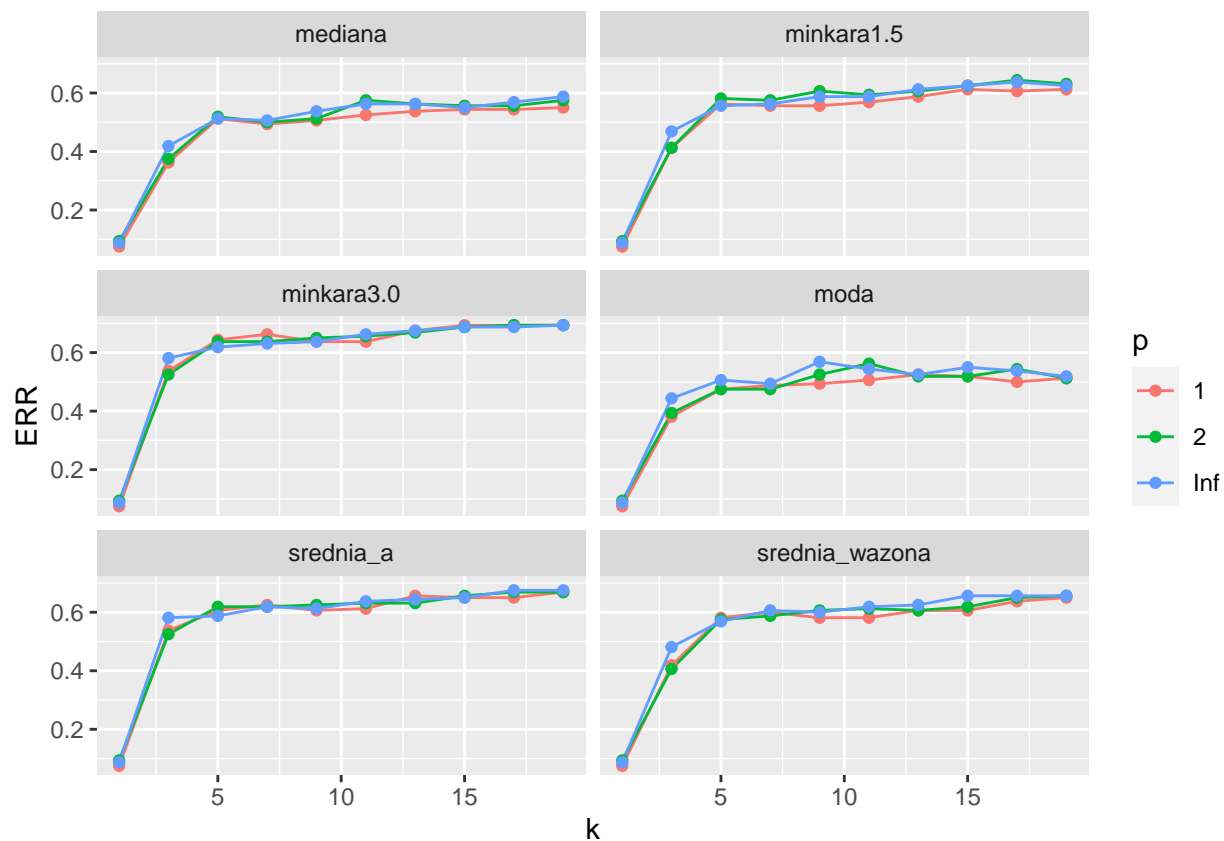
name	ERR	MAD	MSE
randomForest	0.1224927	0.1504382	0.2417722
polr	0.4286595	0.5205777	0.7453749
svm	0.4335281	0.5457644	0.8722817

Zbiór danych *auto_riskness*

Poniżej znajdują się wykresy obrazujące następujące dane:

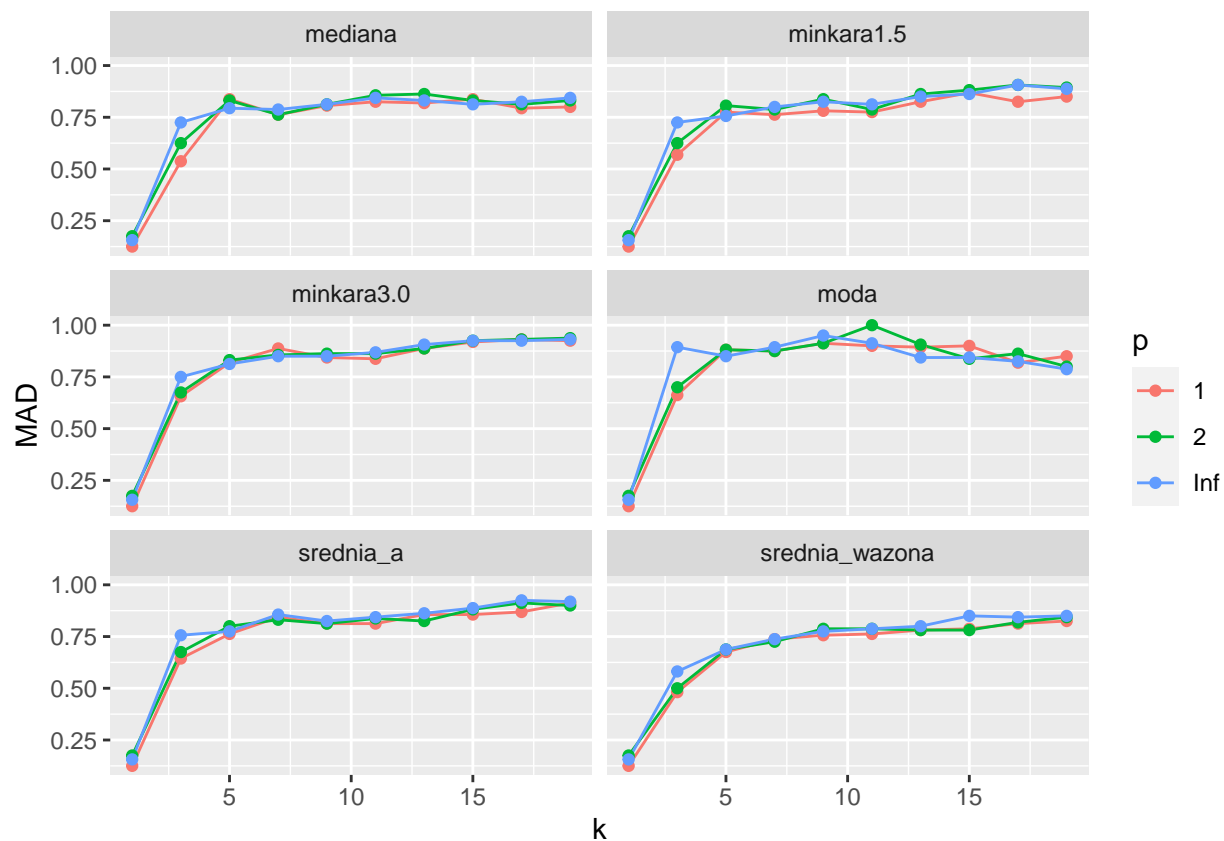
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Wyjątkiem od tej reguły jest tu użycie funkcji agregującej minkara 1.5, która to przy użyciu metryki L_∞ , zaczyna maleć dla $k > 5$. Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



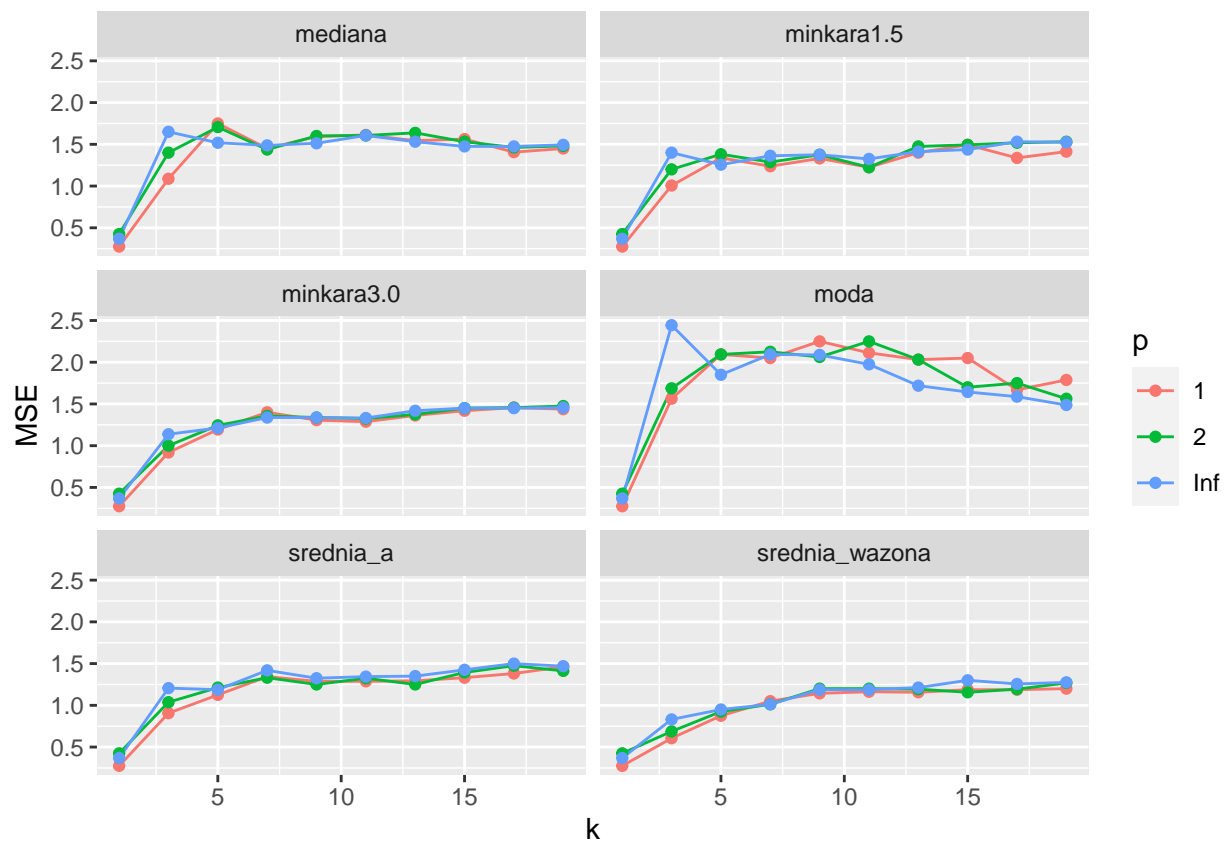
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Najmniejszy błąd osiągnięty jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3, choć w przypadku funkcji *srednia_wazona* jest on zdecydowanie łagodniejszy. Nieznaczne różnice również powodują użyte metryki.



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem średniokwadratowym w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. W tym przypadku *moda* cechuje się największym błędem średniokwadratowym. tu również widzimy pewne rozbieżności między błędami w przypadku użycia różnych metryk. Metryka L_∞ daje największe błędy w tym przypadku.



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru funkcja *randomForest* cechuje się najmniejszymi błędami. Ciekawe jest że dla tej funkcji ERR, MAD i MSE są równe. Następnie funkcja *svm* ma wszystkie błędy o rząd większe od funkcji *randomForest*. Najgorsze wyniki w tym przypadku daje funkcja *polr*

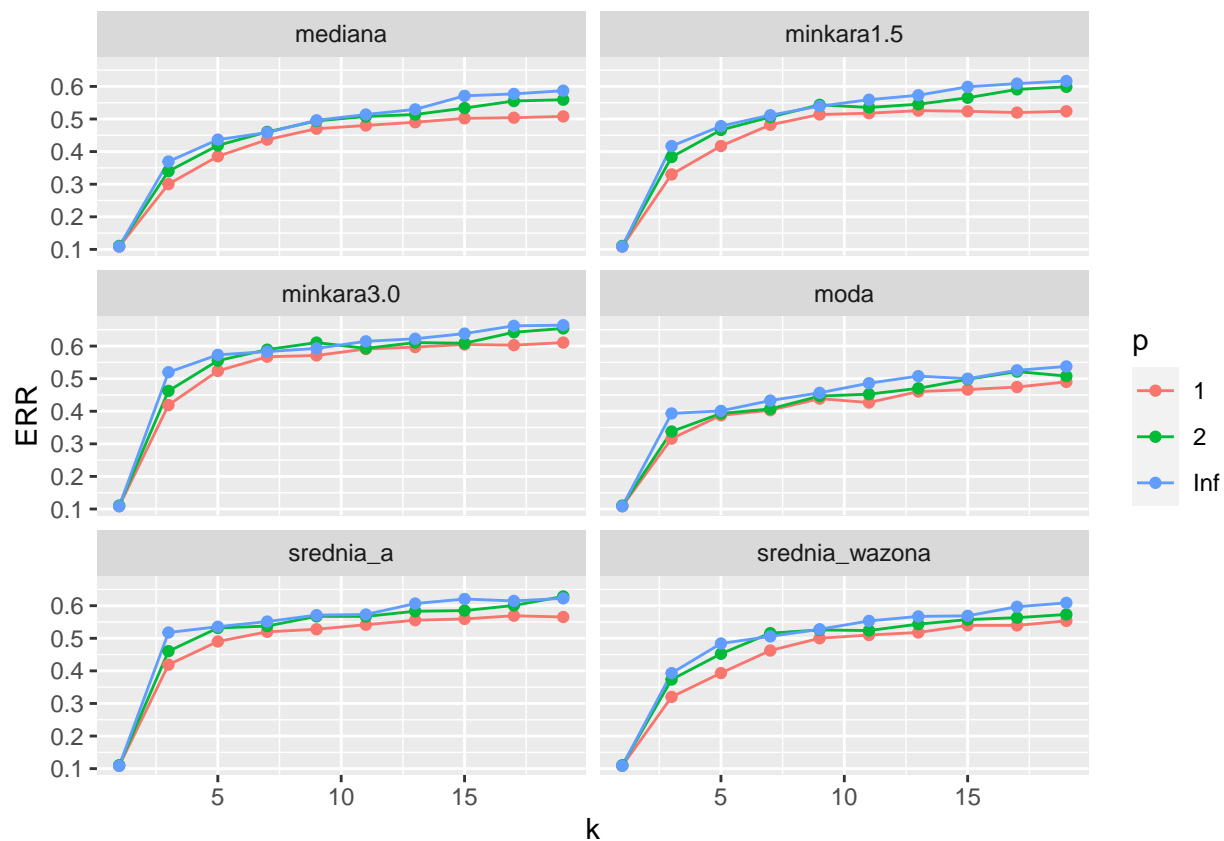
name	ERR	MAD	MSE
randomForest	0.02500	0.02500	0.02500
polr	0.48750	0.55625	0.70625
svm	0.24375	0.36875	0.64375

Zbiór danych *bostonhousing*

Poniżej znajdują się wykresy obrazujące następujące dane:

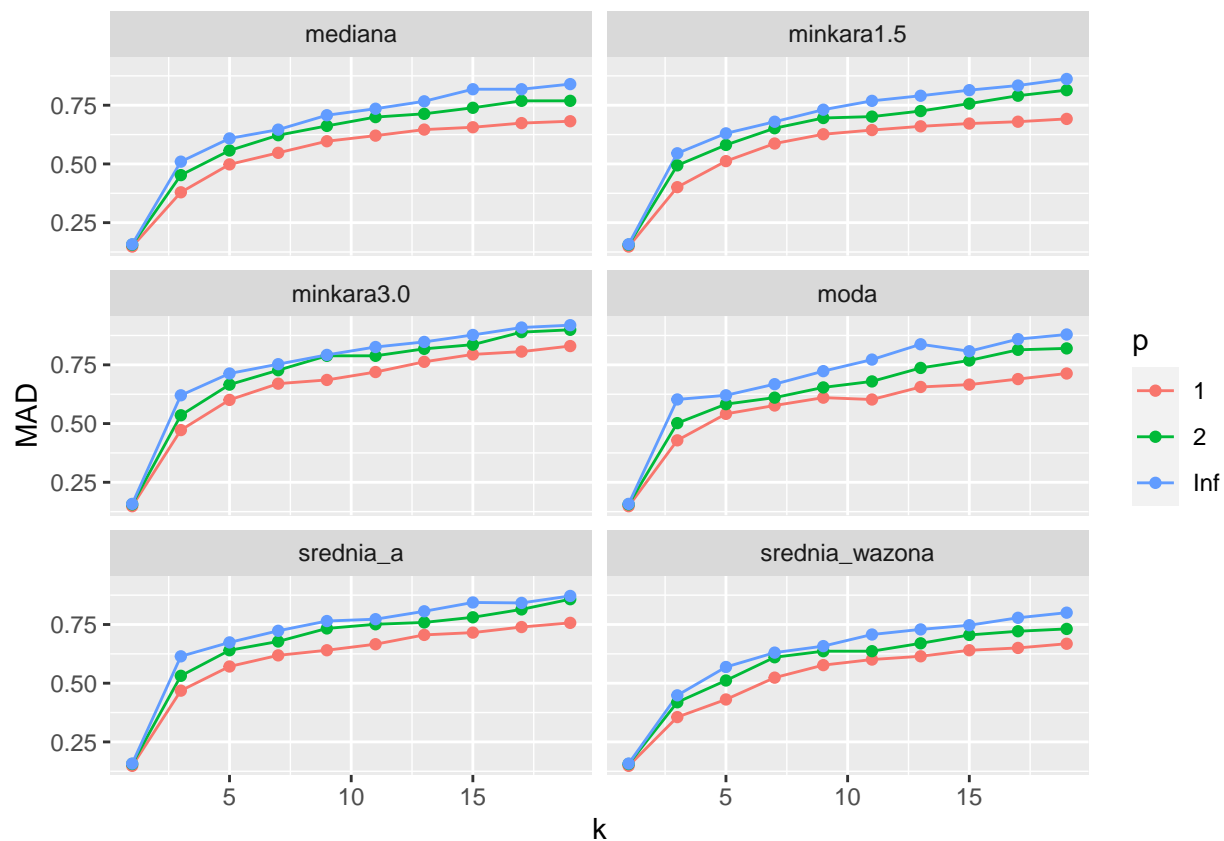
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że błąd ERR zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. Widać też że dla niektórych k zauważalnie zmniejsza się błąd dla metryki L1



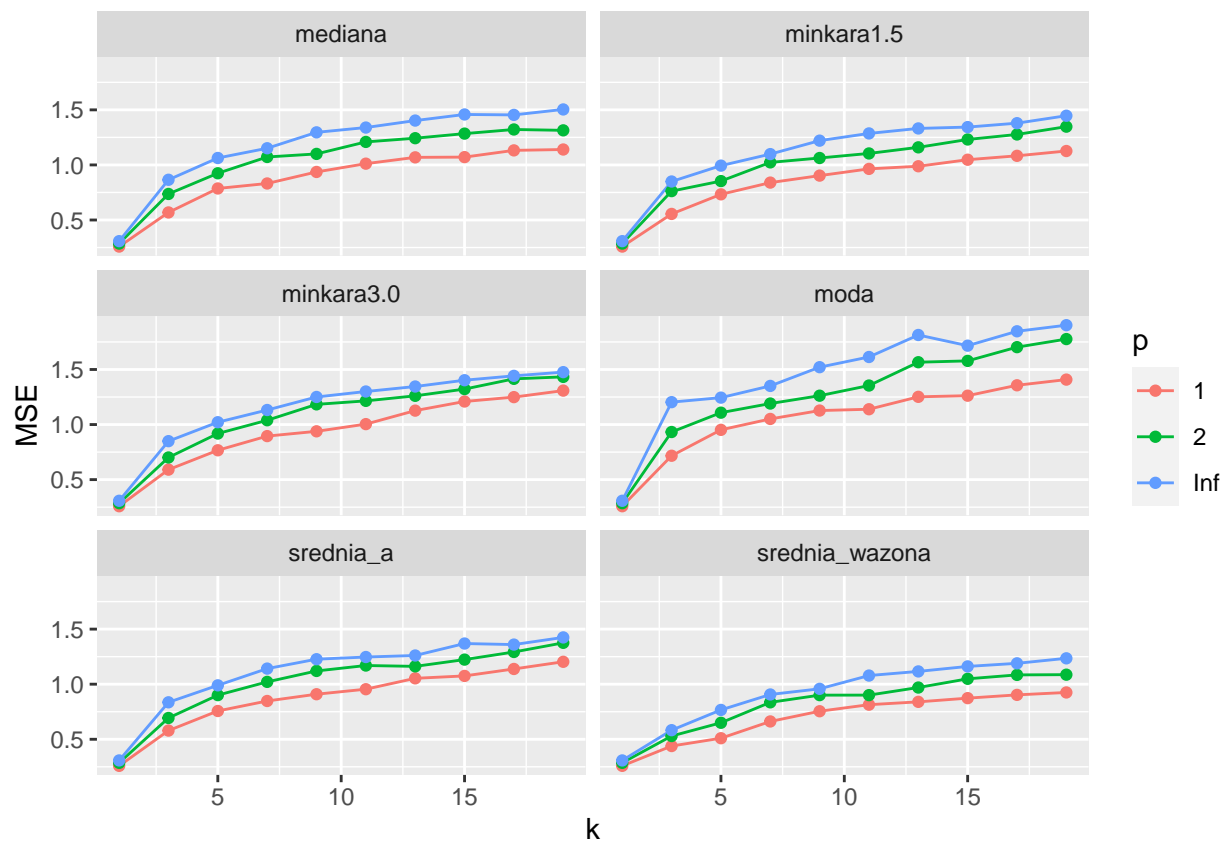
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem bezwzględnych w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . W tym przypadku widać że można zaobserwować niższe błędy dla metryki L1 i odrobinę wyższe dla metryki L2 i najwyższe dla L_∞ . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem średniokwadratowym i najłagodniejszym wzrostem w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. W tym przypadku widać że można zaobserwować niższe błędy dla metryki L1, trochę większe dla L2 i największe dla L_∞ . W tym przypadku moda cechuje się największym błędem średniokwadratowym.



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru danych można zaobserwować że najlepiej radzi sobie funkcja *randomForest* z najmniejszymi wszystkimi błędami. Błąd ERR tej funkcji jest porównywalny z błędem 1-nn. Błędy ERR pozostałych funkcji są mniejsze od błędów knn dla $k = 3$ oraz większe niż dla 1-nn. Wszystkie te funkcje cechują zauważalnie mniejszymi błędami MAD i MSE.

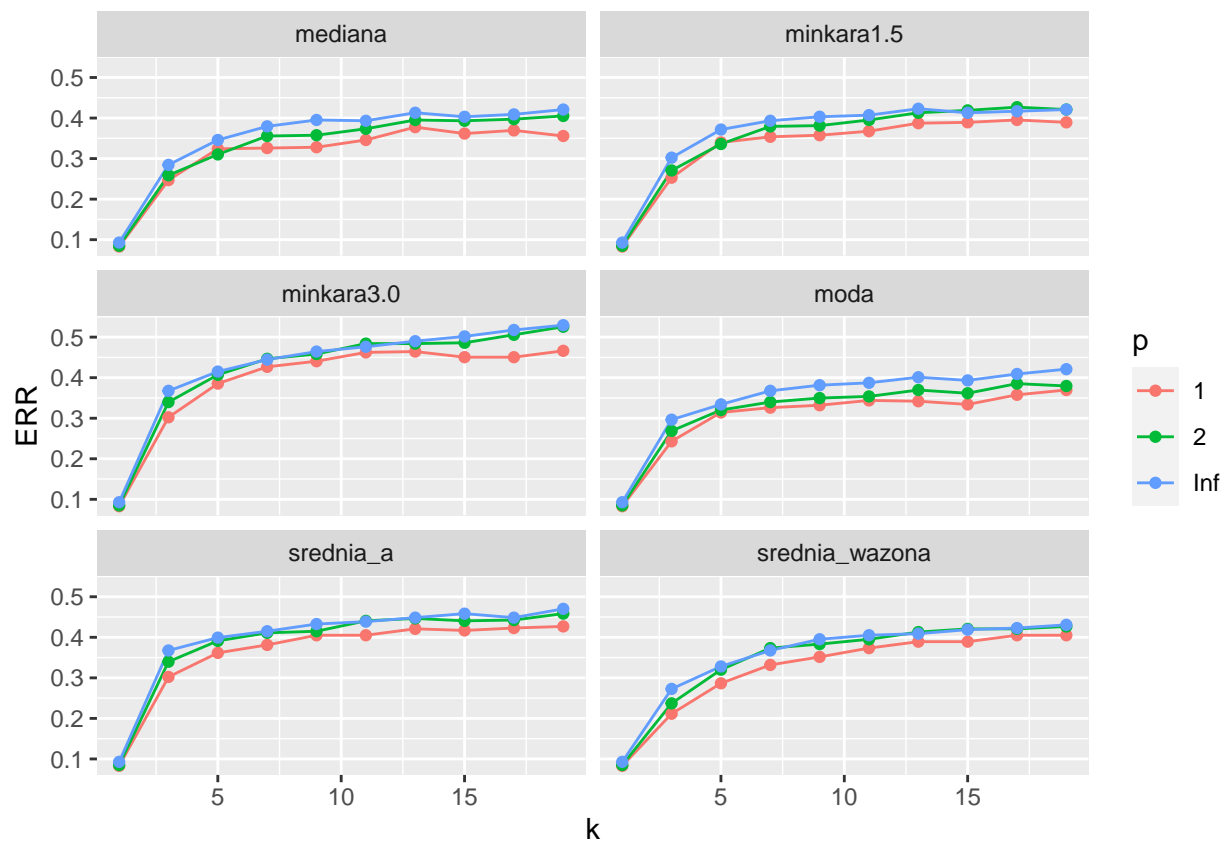
name	ERR	MAD	MSE
randomForest	0.06125024	0.06913221	0.08885653
polr	0.37542225	0.44255484	0.62026791
svm	0.34377791	0.42083091	0.63026597

Zbiór danych *bostonhousing_ord*

Poniżej znajdują się wykresy obrazujące następujące dane:

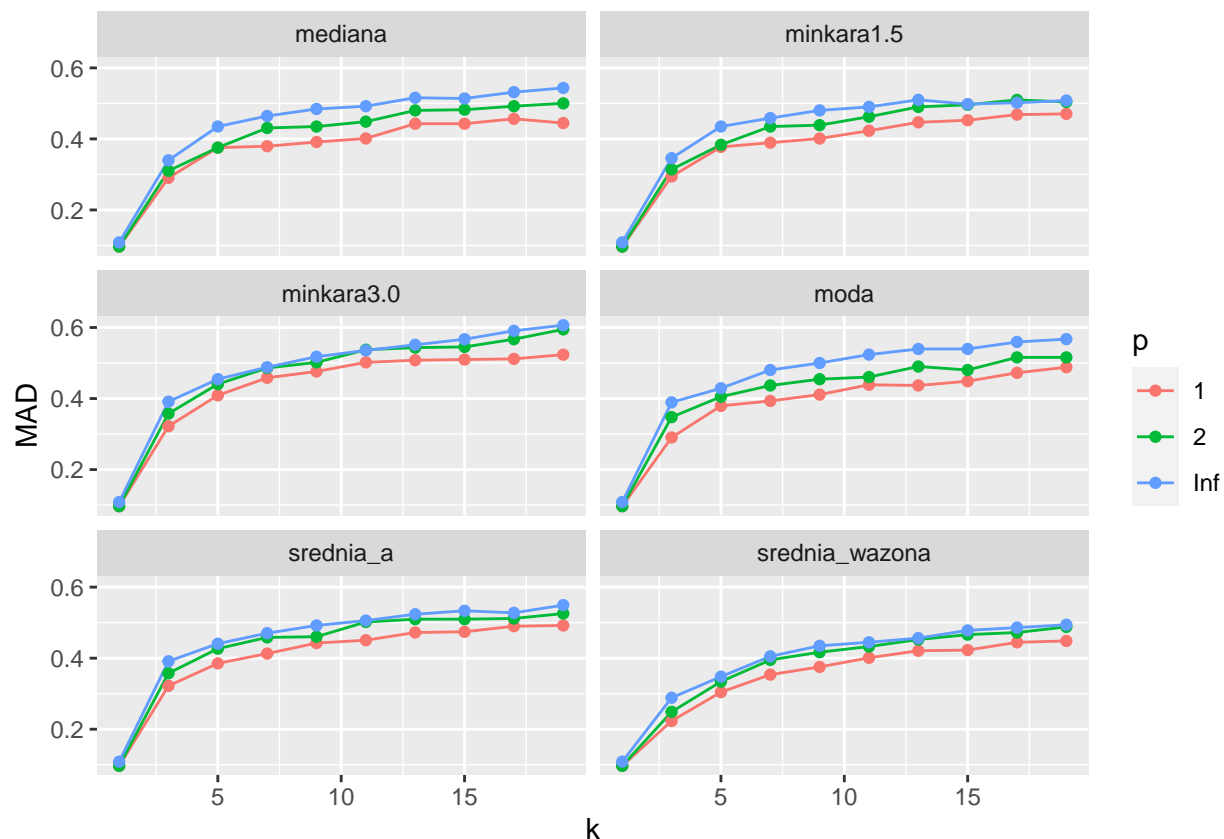
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że błąd ERR zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. Widać też że dla niektórych k zauważalnie zmniejsza się błąd dla metryki L1



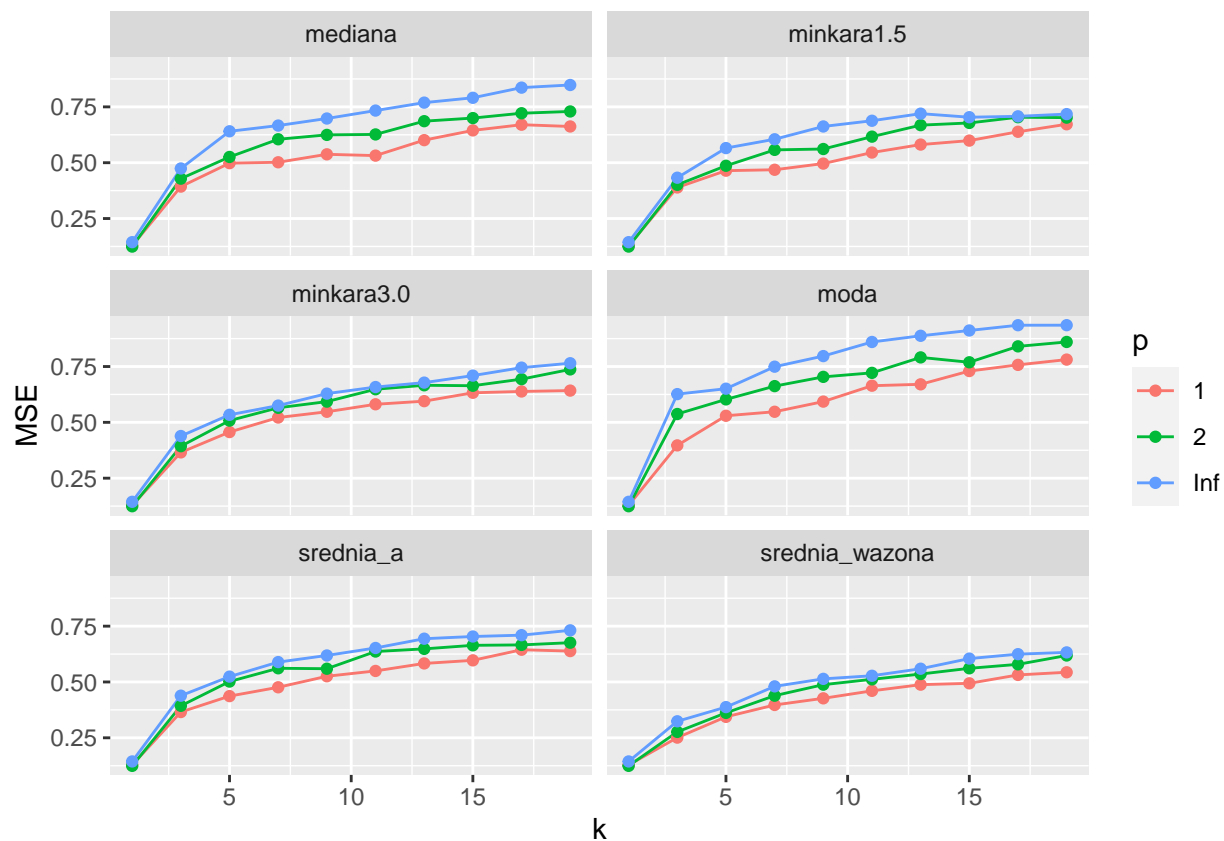
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem bezwzględnych w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . W tym przypadku widać że można zaobserwować niższe błędy dla metryki L1 i odrobinę wyższe dla metryki L2 i najwyższe dla L_∞ . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem średniokwadratowym i najłagodniejszym wzrostem w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. W tym przypadku widać że można zaobserwować niższe błędy dla metryki L1, trochę większe dla L2 i największe dla L_∞ . W tym przypadku *moda* cechuje się największym błędem średniokwadratowym.



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru danych można zaobserwować że najlepiej radzi sobie funkcja *randomForest* z najmniejszymi wszystkimi błędami. Błąd ERR tej funkcji jest porównywalny z błędem 1-nn. Błędy ERR pozostałych funkcji są porównywalne do błędów knn dla $k = 3$. Wszystkie te funkcje cechują zauważalnie mniejszymi błędami MAD i MSE.

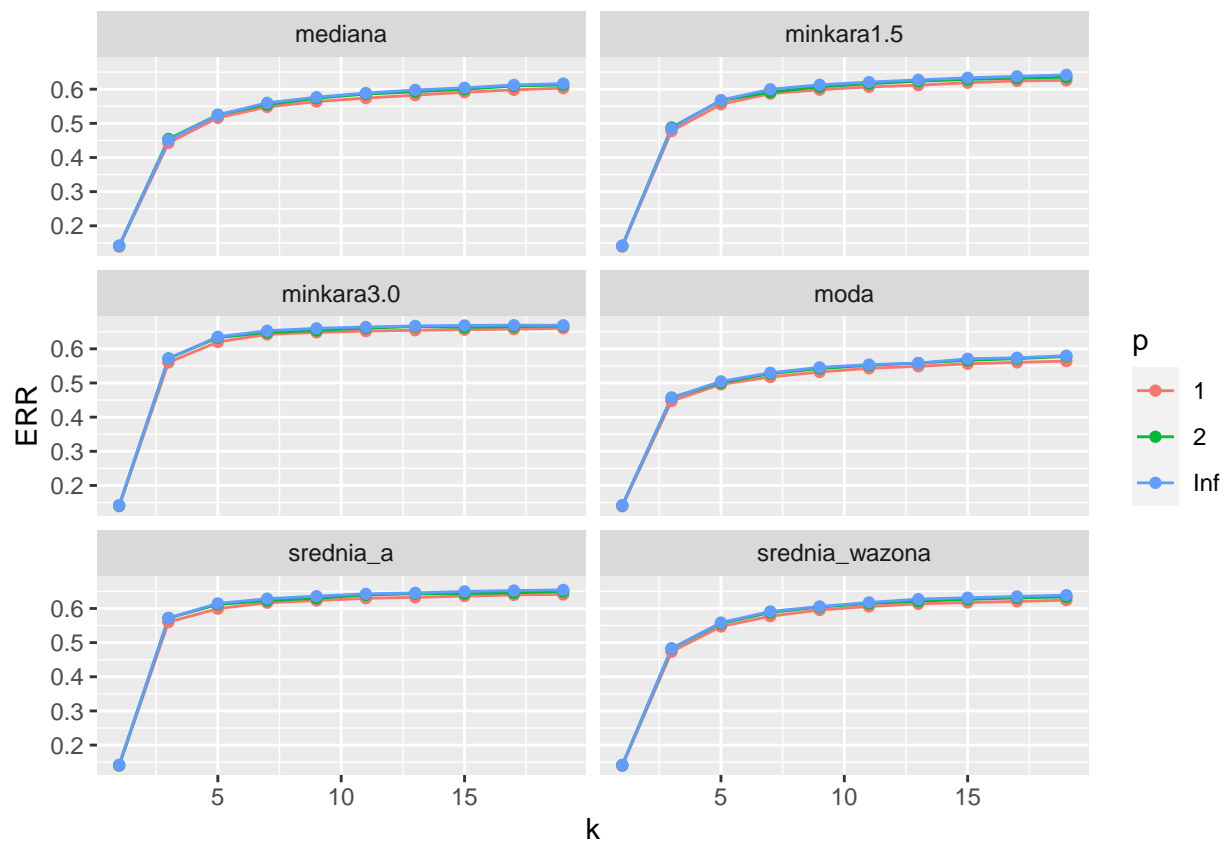
name	ERR	MAD	MSE
randomForest	0.05730926	0.06523005	0.08899243
polr	0.25892060	0.28466317	0.35198990
svm	0.23913803	0.25894001	0.31438556

Zbiór danych *californiahousing*

Poniżej znajdują się wykresy obrazujące następujące dane:

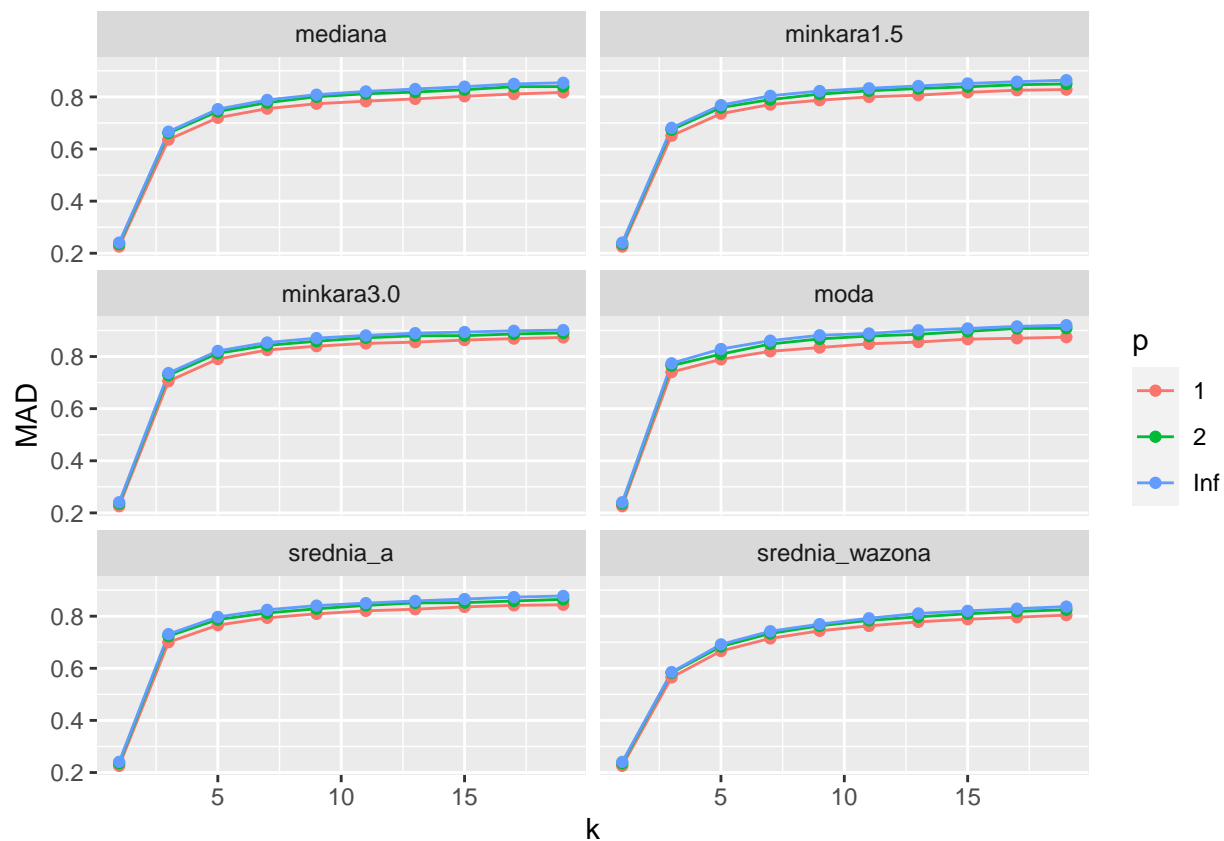
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że błąd ERR zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



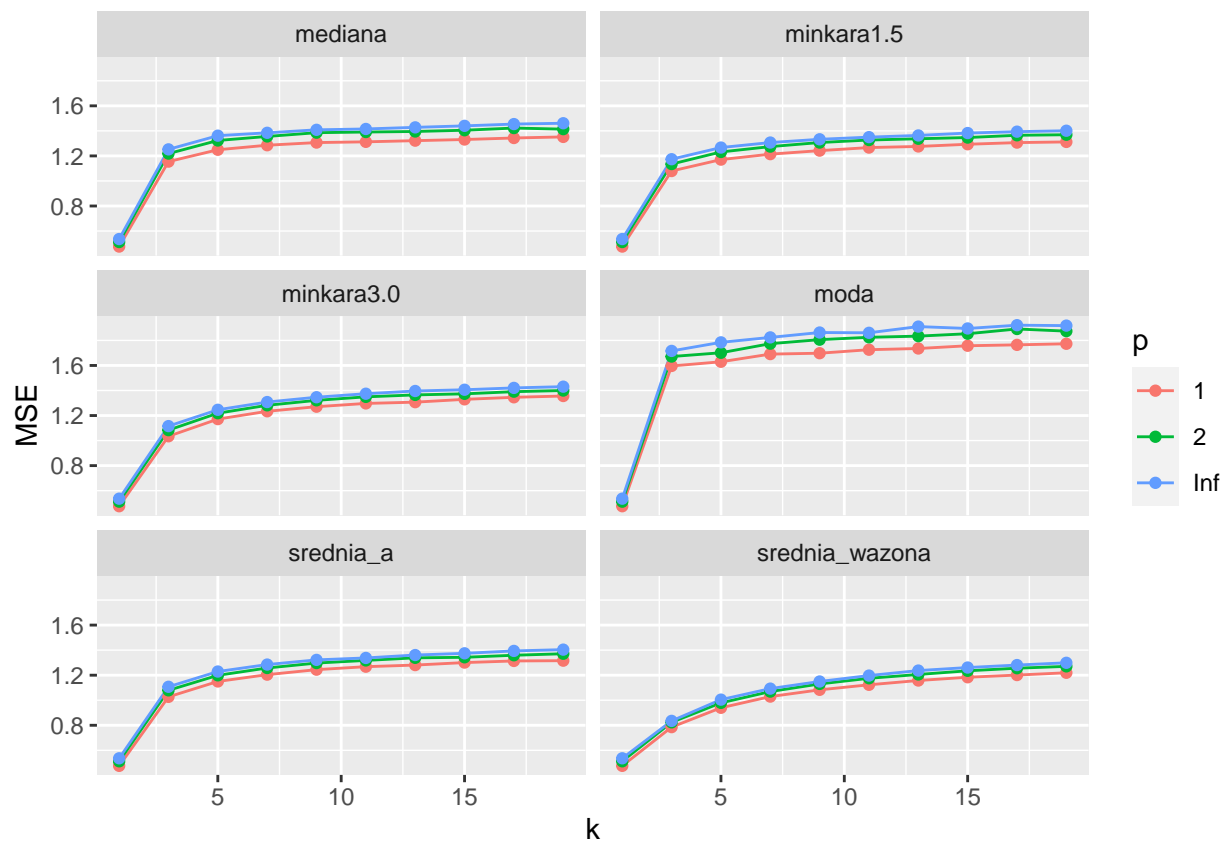
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem bezwzględnych w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem średniokwadratowym w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. W tym przypadku widać że minimalnie można zaobserwować niższe błędy dla metryki L1. W tym przypadku *moda* cechuje się największym błędem średniokwadratowym.



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru danych można zaobserwować że najlepiej radzi sobie funkcja *randomForest* z najmniejszymi wszystkimi błędami. Błąd ERR tej funkcji jest porównywalny z błędem 1-nn. Błędy funkcji *svm* są mniejsze niż błędy knn dla $k = 3$ oraz większe niż błędy 1-nn. Niestety użycie funkcji *polr* było niemożliwe bez usuwania dodatkowych oprócz liniowo zależnych kolumn ze zbioru danych.

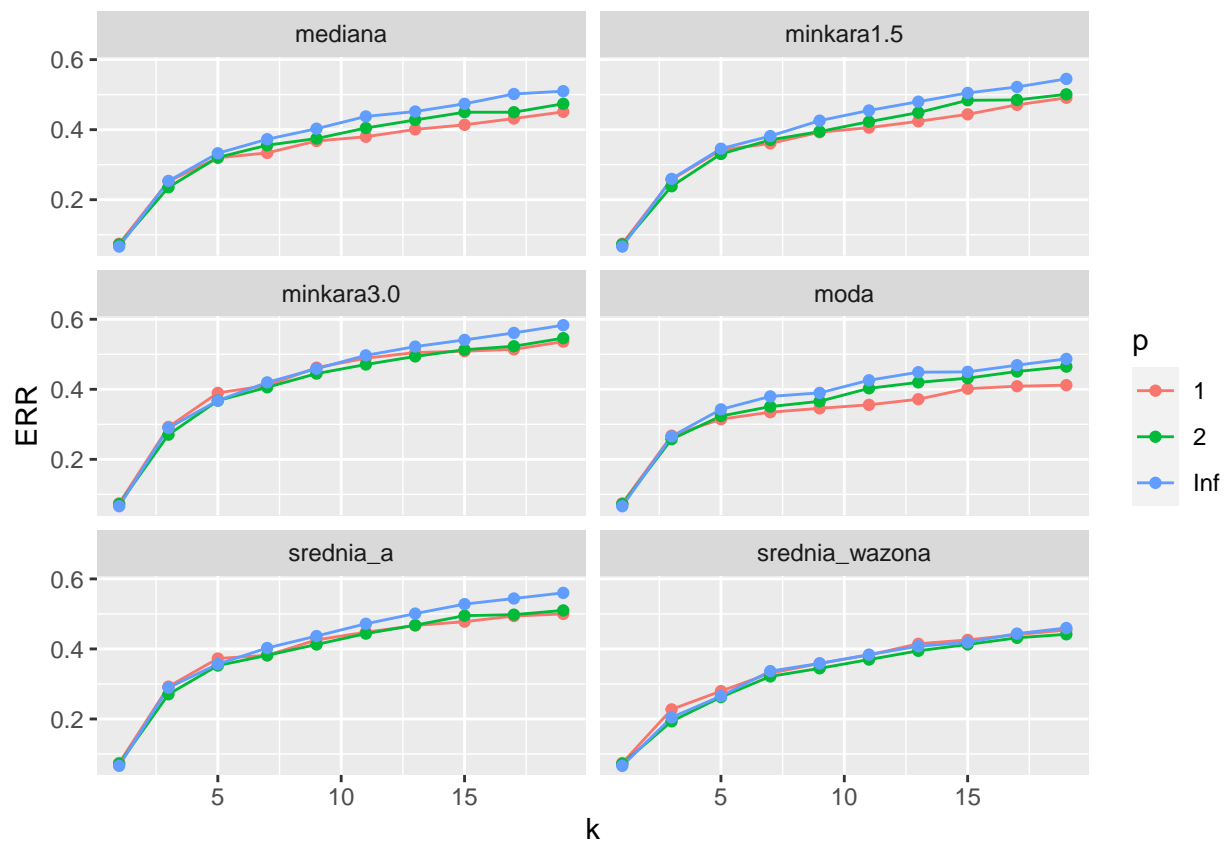
name	ERR	MAD	MSE
randomForest	0.06463178	0.07679264	0.1078973
svm	0.38062016	0.47480620	0.7156008

Zbiór danych *cement_strength*

Poniżej znajdują się wykresy obrazujące następujące dane:

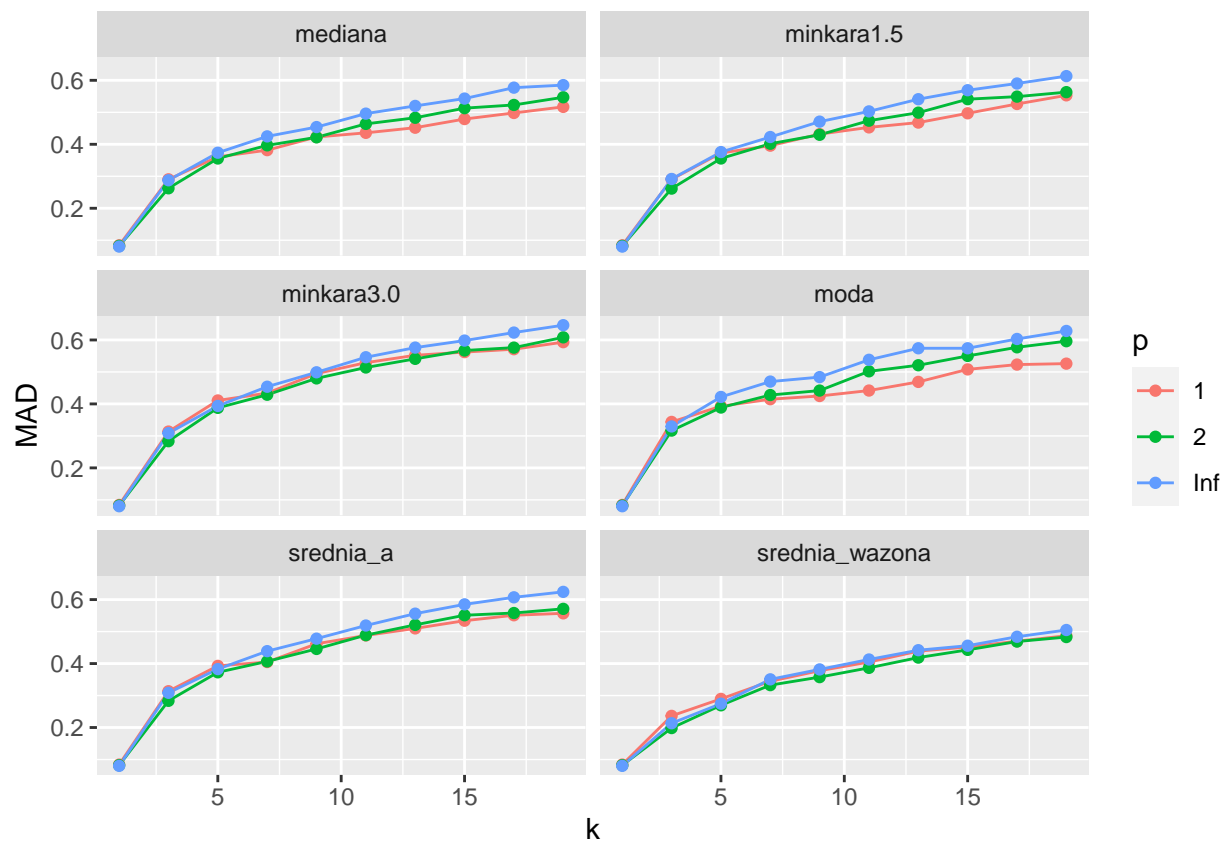
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że błąd ERR zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. Najgorsze wyniki zostały uzyskane dla metryki L_∞ .



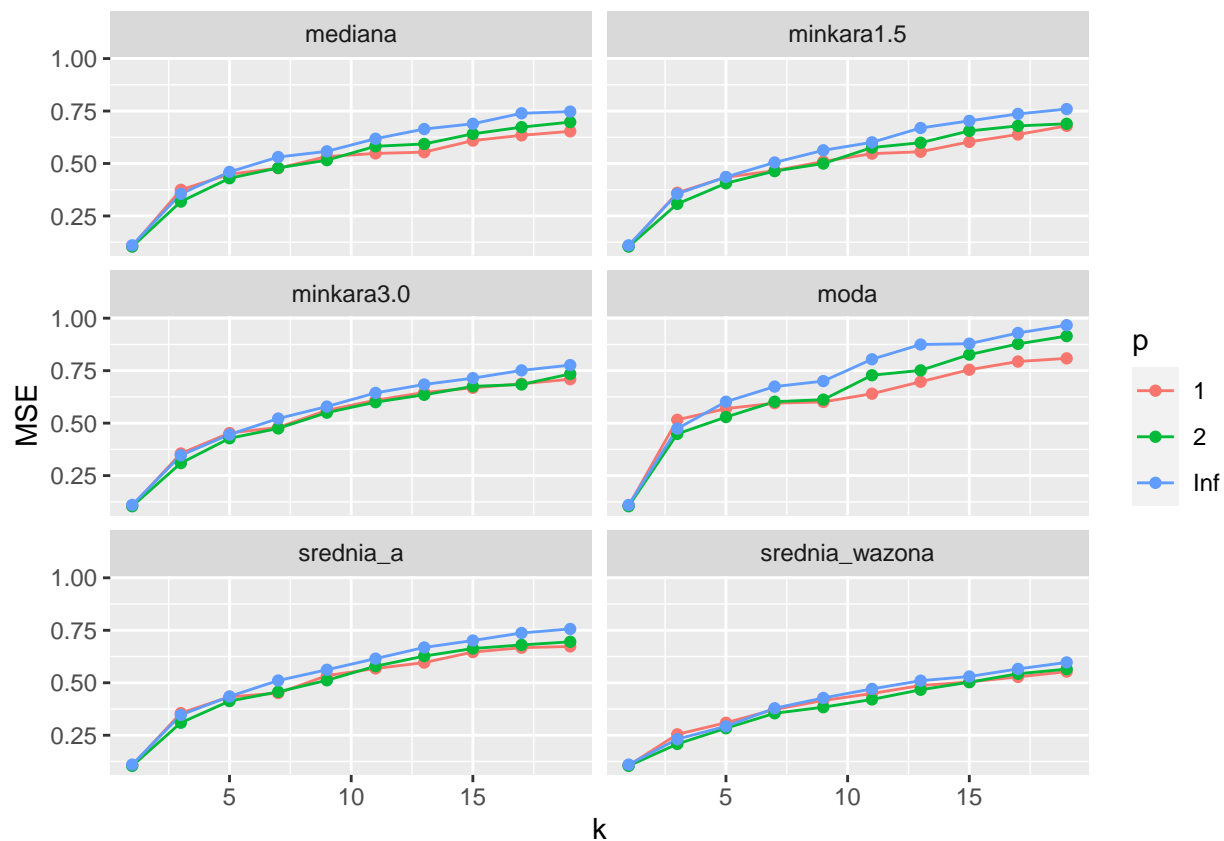
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem bezwzględnych w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. Najgorsze wyniki zostały uzyskane dla metryki L_∞ .



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem średniokwadratowym w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. W tym przypadku widać minimalnie niższe błędy dla metryki L1 i L2. W tym przypadku moda cechuje się największym błędem średniokwadratowym. Najgorsze wyniki zostały uzyskane dla metryki L_∞ . Błędy dla tego zbioru rosną w bardziej płaski sposób niż w innych funkcjach.



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru danych można zaobserwować że najlepiej radzi sobie funkcja *randomForest* z najmniejszymi wszystkimi błędami. Błąd ERR tej funkcji jest porównywalny z błędem 1-nn. Błędy funkcji *svm* są porównywalne z błędem knn dla $k = 3$. Niestety użycie funkcji *polr* dało najgorsze wyniki porównywalne z knn dla dużego k .

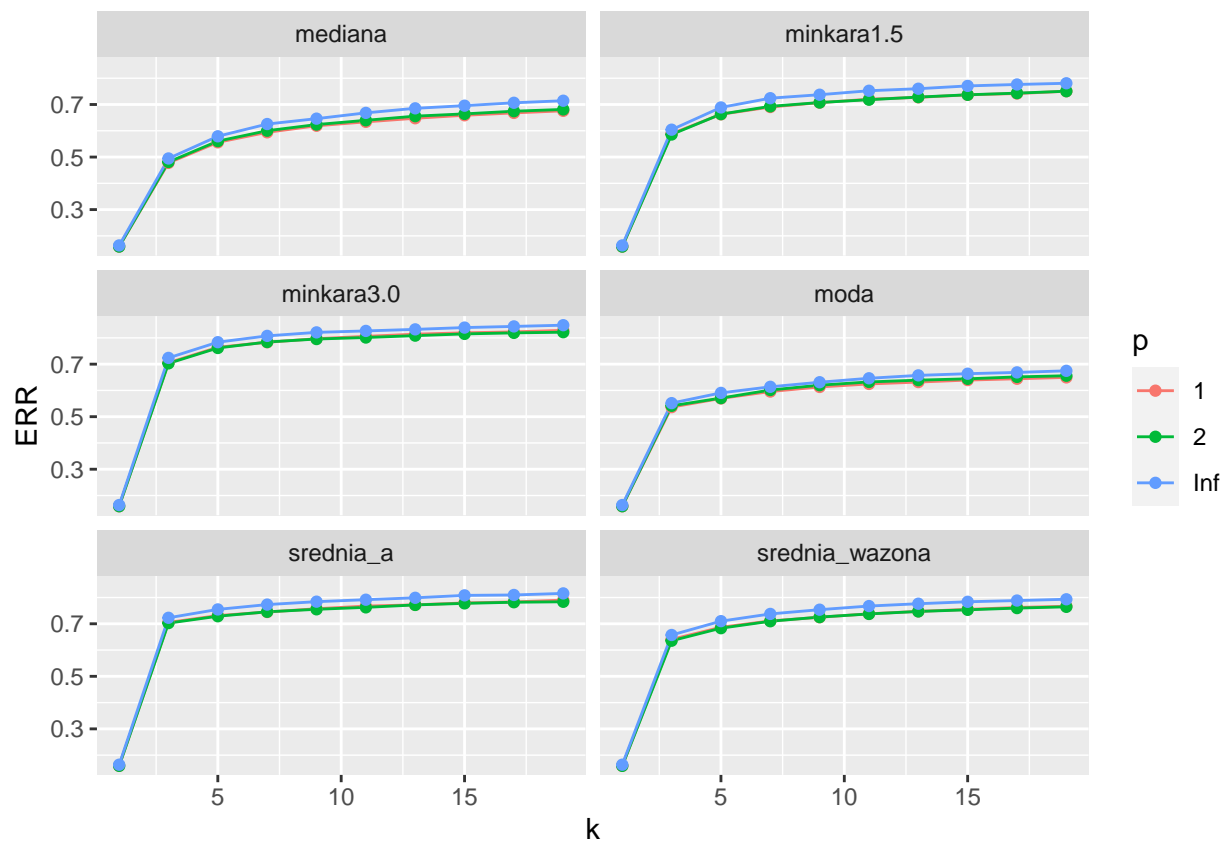
name	ERR	MAD	MSE
randomForest	0.06613568	0.07115075	0.0811809
polr	0.51607538	0.59626131	0.7666633
svm	0.32864322	0.37776382	0.4860352

Zbiór danych *fireman_example*

Poniżej znajdują się wykresy obrazujące następujące dane:

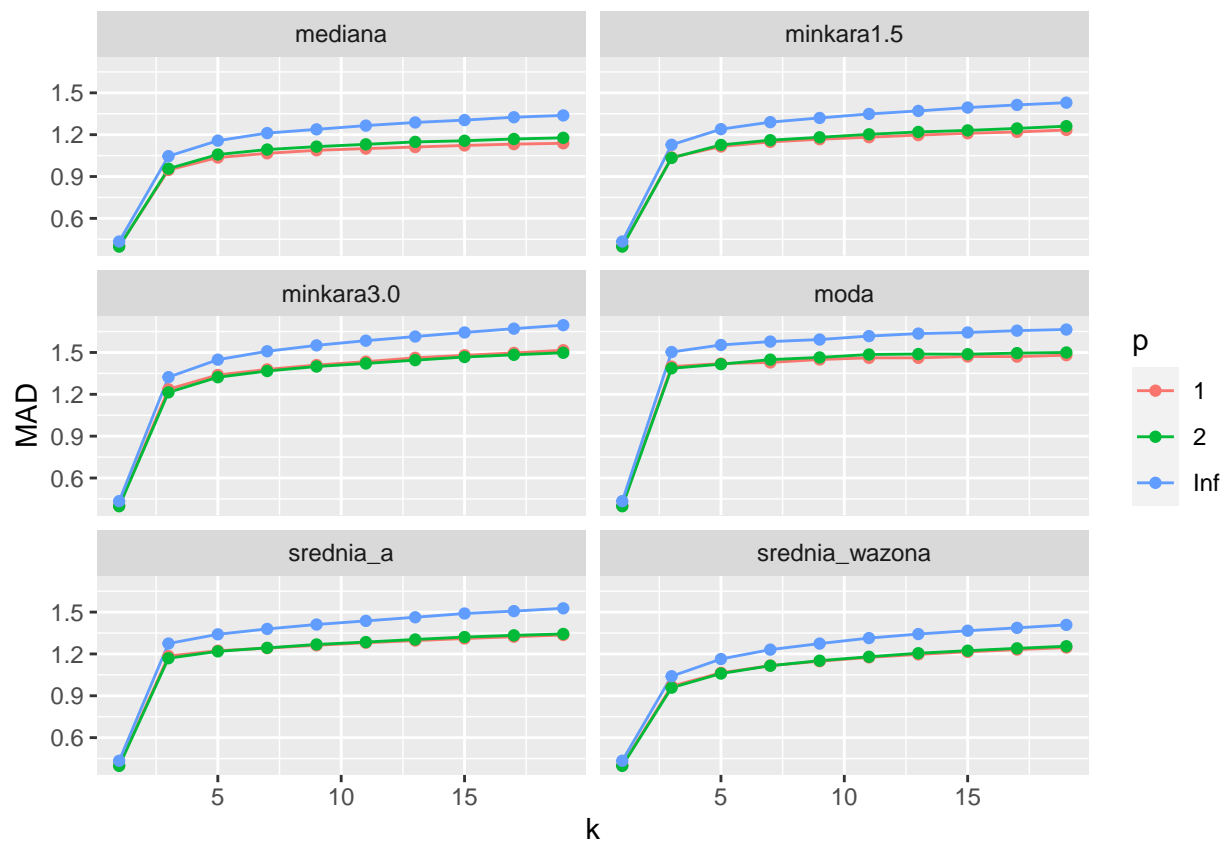
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że błąd ERR zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



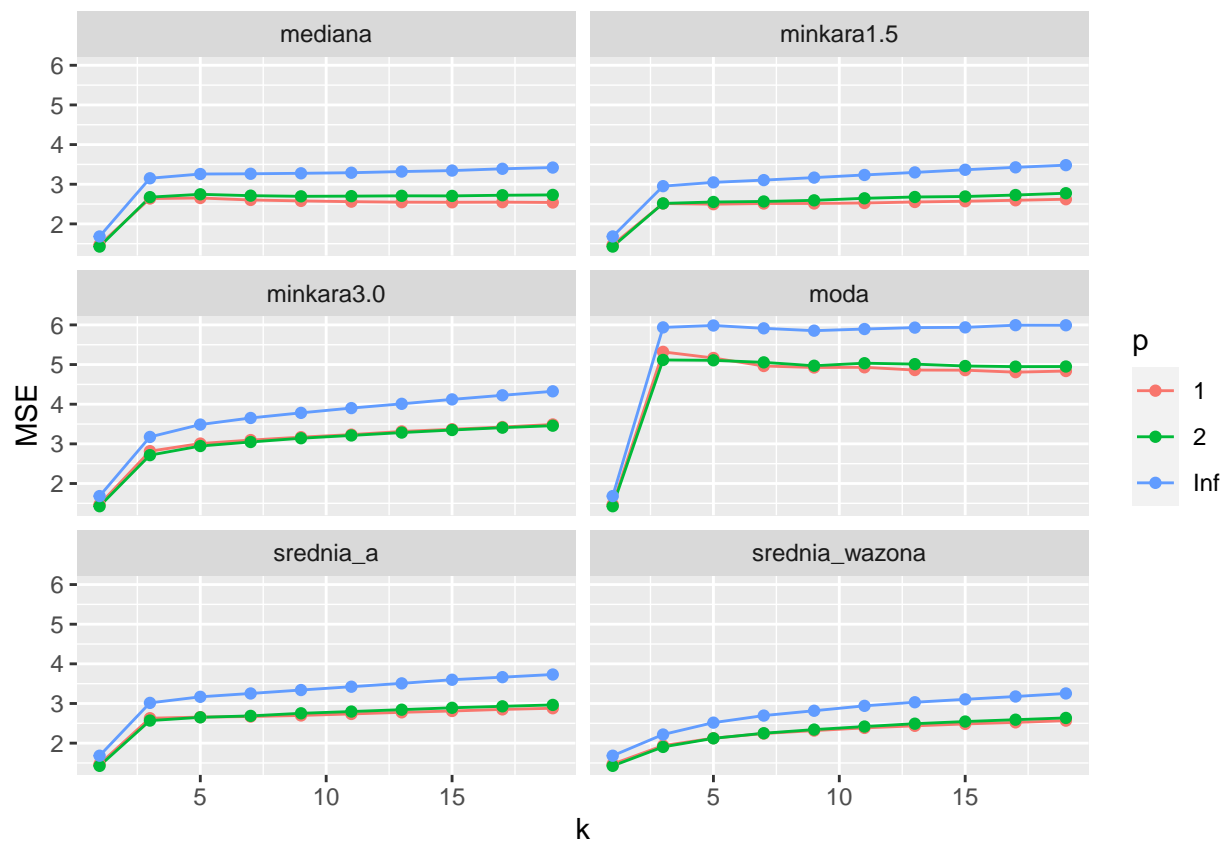
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem bezwzględnym w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. Najgorsze wyniki zostały uzyskane dla metryki L_∞ .



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem średniokwadratowym w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. W tym przypadku widać że można zaobserwować niższe błędy dla metryki L1 i L2. W tym przypadku moda cechuje się największym błędem średniokwadratowym. Najgorsze wyniki zostały uzyskane dla metryki L_∞ .



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru danych można zaobserwować że najlepiej radzi sobie funkcja *randomForest* z najmniejszymi wszystkimi błędami. Błąd ERR tej funkcji jest porównywalny z błędem 1-nn. Błędy funkcji *svm* są mniejsze niż błędy knn dla $k = 3$ oraz większe niż błędy 1-nn. Niestety użycie funkcji *polr* dało najgorsze wyniki porównywalne z knn dla $k \geq 3$ dla funkcji moda.

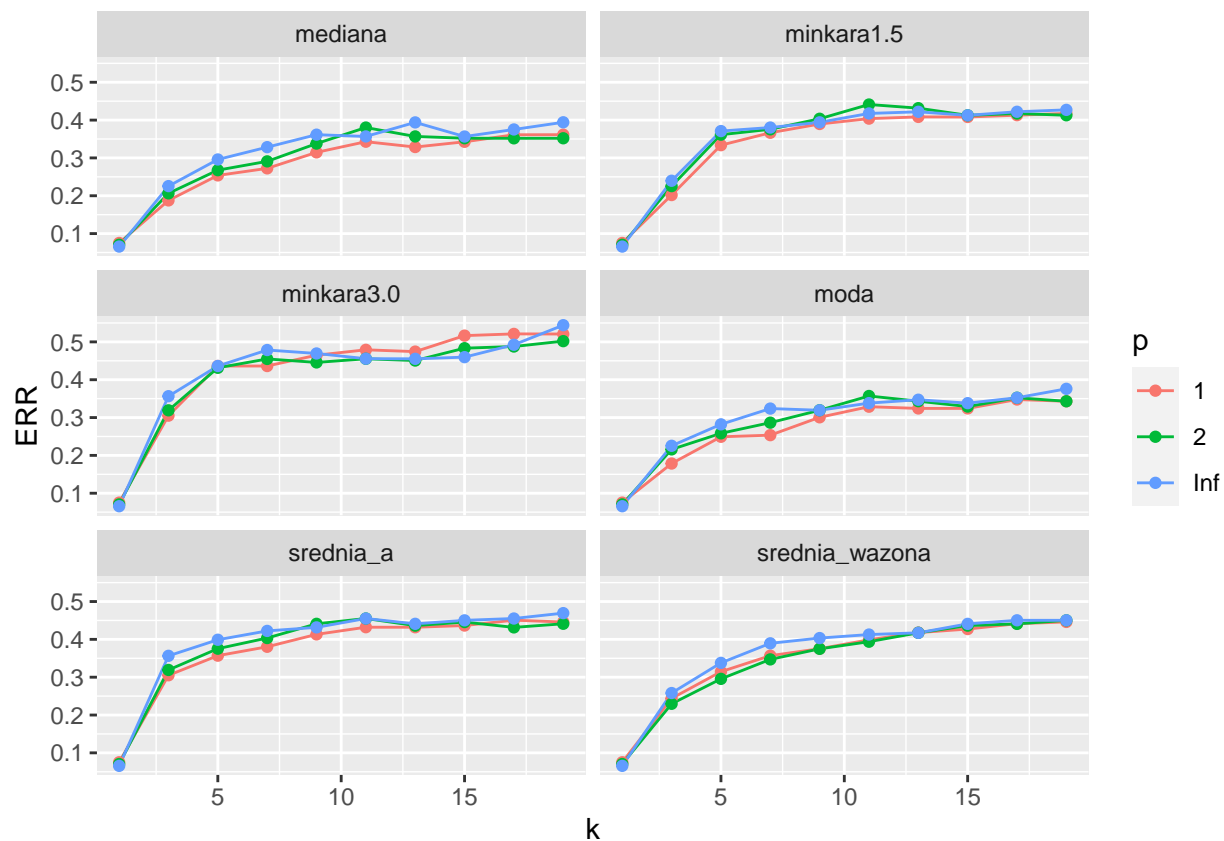
name	ERR	MAD	MSE
randomForest	0.1339287	0.2293957	0.5373327
polr	0.7626815	1.8284436	6.6958853
svm	0.5433184	0.8510845	1.7622409

Zbiór danych *glass*

Poniżej znajdują się wykresy obrazujące następujące dane:

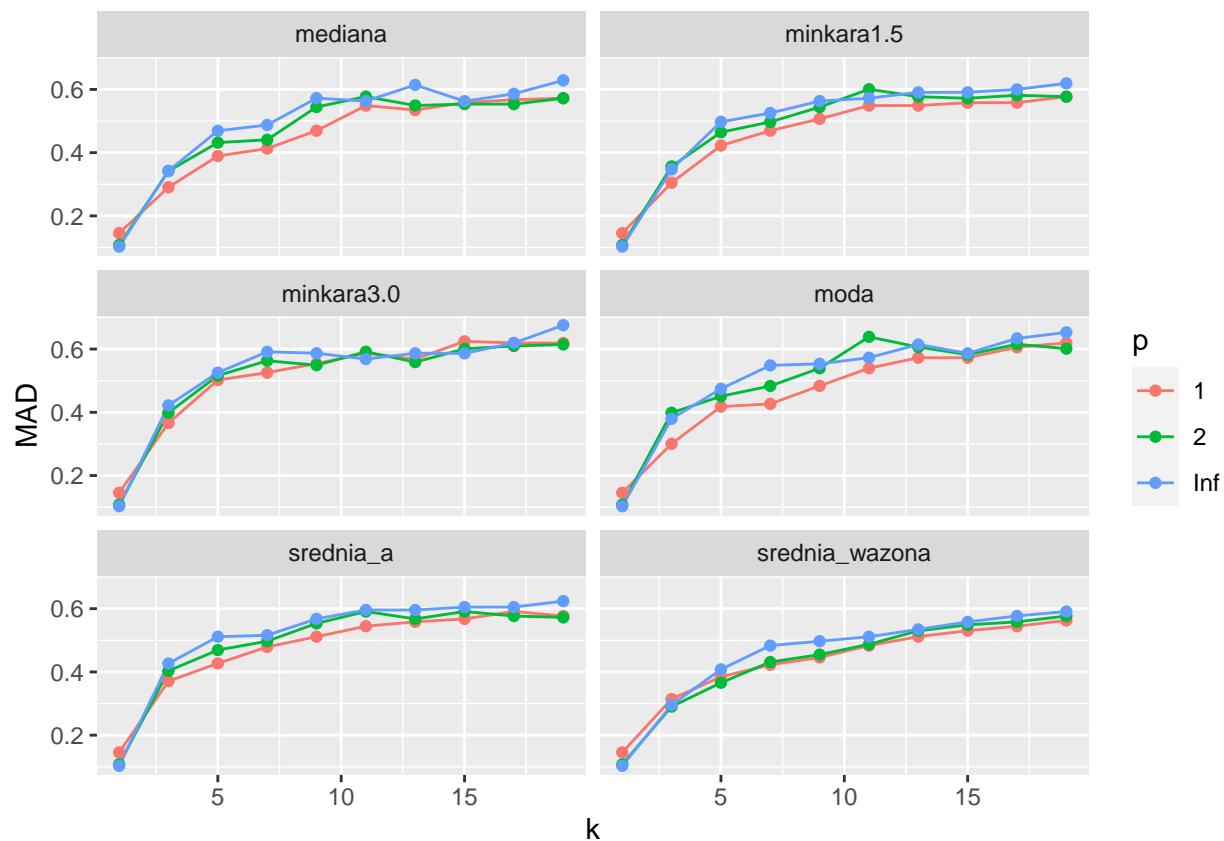
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że błąd ERR zachowuje się w podobny sposób dla wszystkich funkcji *minkara1.5*, *moda*, *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 5. Najgorsze wyniki zostały uzyskane dla metryki L_∞ . Dla funkcji *minkara3.0* i *srednia_a* wykresy są strome dla k od 1 do 3, a następnie już się spłaszczają.



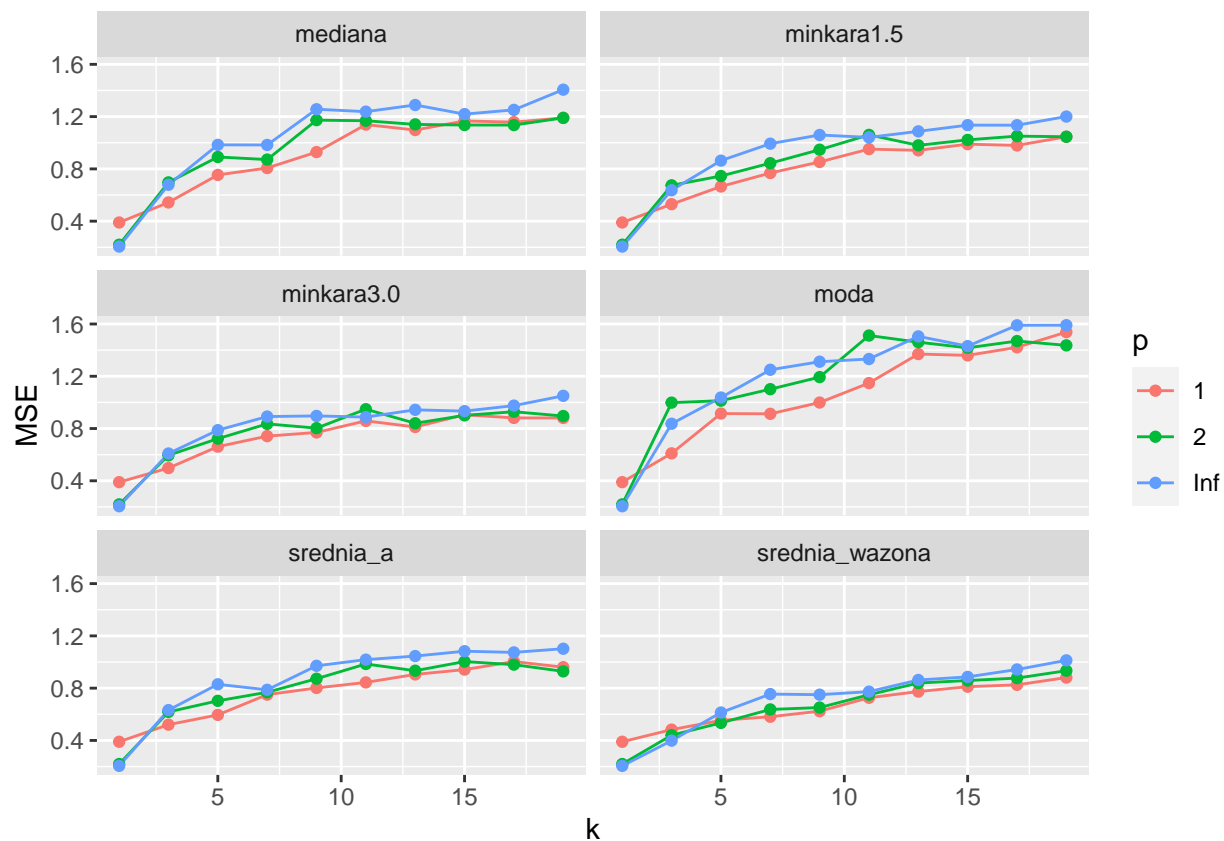
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem bezwzględnych w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. Najgorsze wyniki zostały uzyskane dla metryki L_∞ .



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem średniokwadratowym w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 5. W tym przypadku widać minimalnie niższe błędy dla metryki L1 i L2. W tym przypadku *moda* cechuje się największym błędem średniokwadratowym. Najgorsze wyniki zostały uzyskane dla metryki L_∞ .



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru danych można zaobserwować że najlepiej radzi sobie funkcja *randomForest* z najmniejszymi wszystkimi błędami. Błąd ERR tej funkcji jest porównywalny z błędem 1-nn. Błędy funkcji *svm* są porównywalne z błędem knn dla $k = 3$. Niestety użycie funkcji *polr* dało najgorsze wyniki porównywalne z knn dla dużego k .

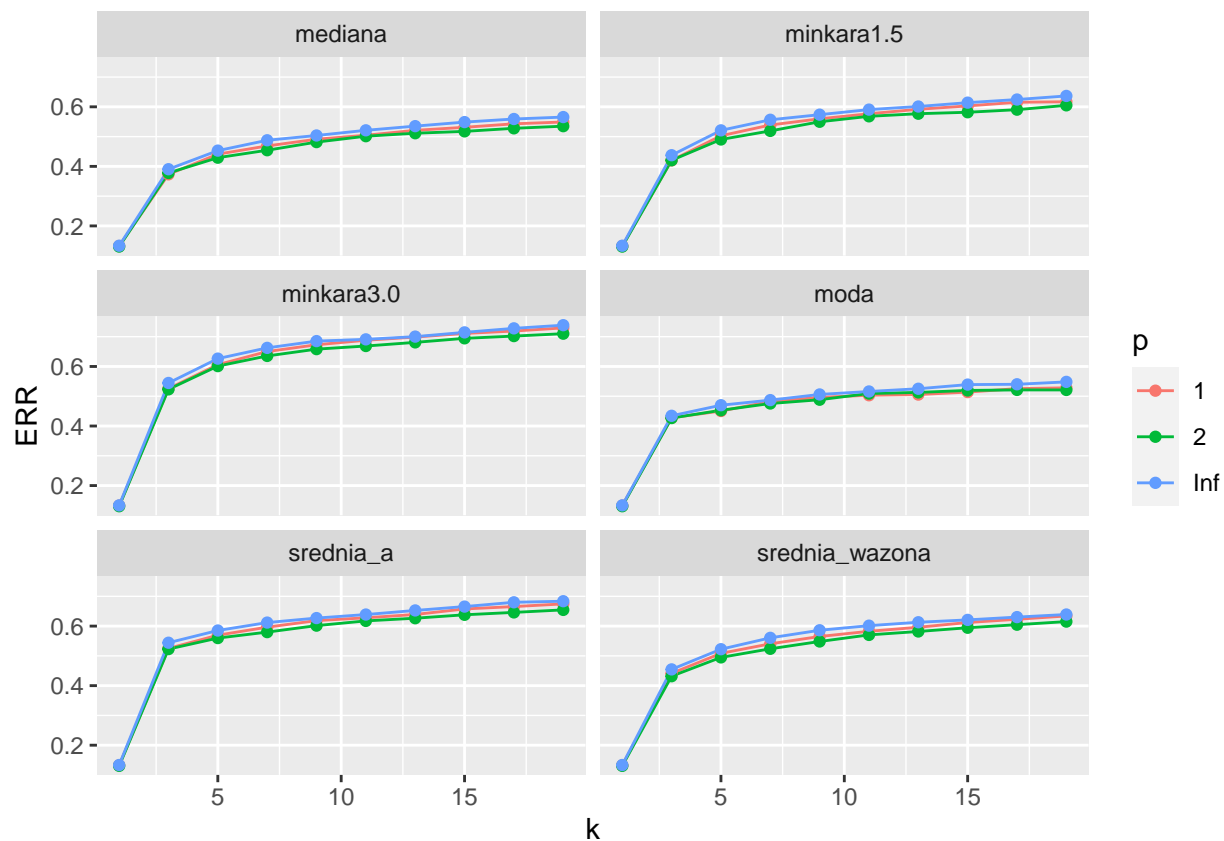
name	ERR	MAD	MSE
randomForest	0.03277962	0.07984496	0.2213732
polr	0.49324474	0.70930233	1.2920266
svm	0.20708749	0.30155039	0.5854928

Zbiór danych *kinematics*

Poniżej znajdują się wykresy obrazujące następujące dane:

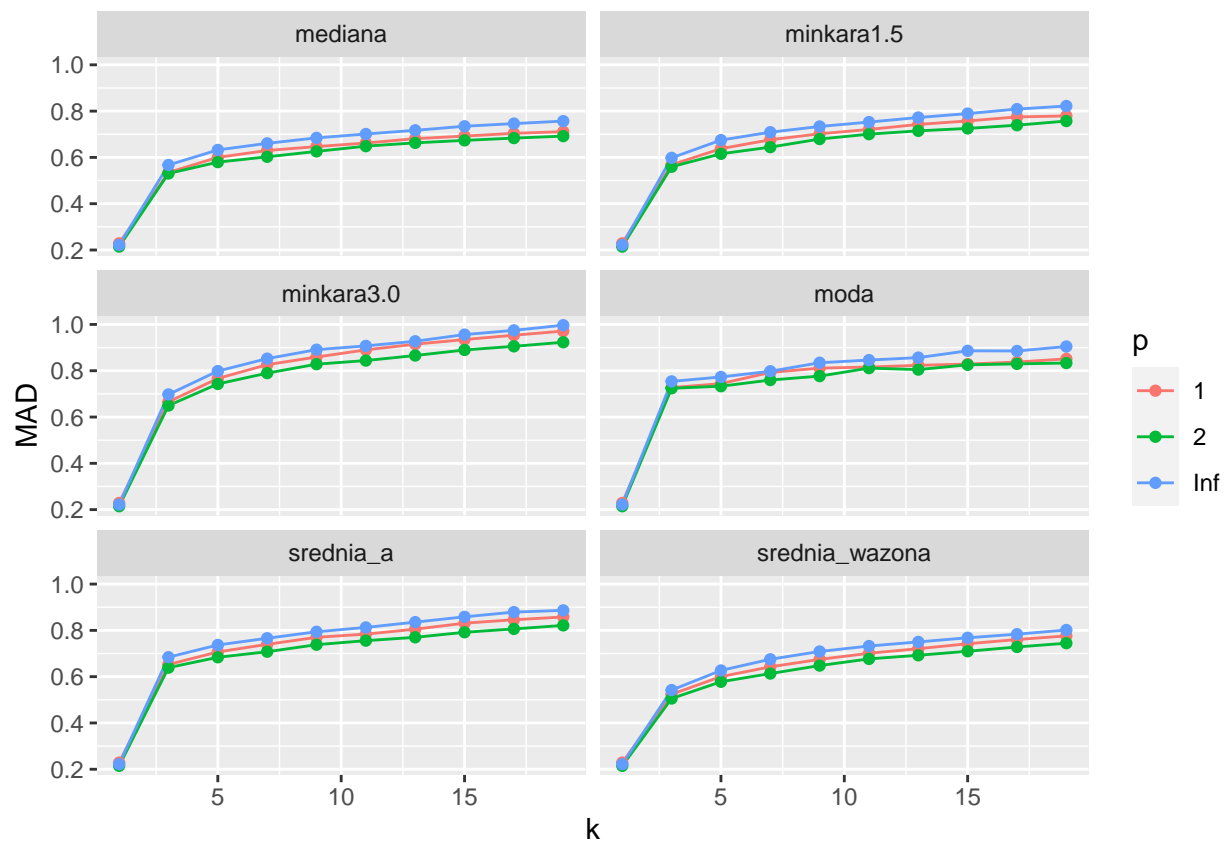
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że błąd ERR zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. Najgorsze wyniki dają funkcje *minkara3.0* i *srednia_a*, a najlepsze *moda*.



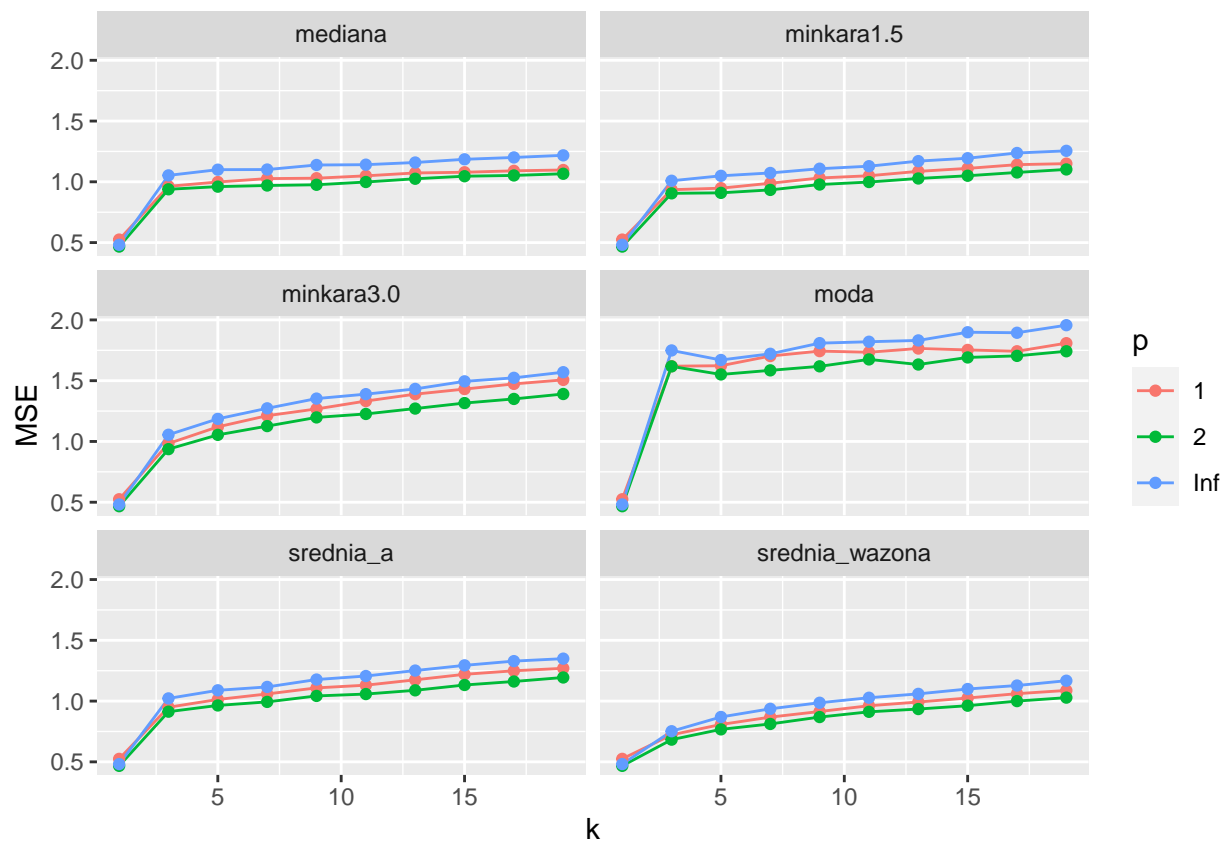
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem bezwzględnych w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. Najgorsze wyniki zostały uzyskane dla metryki L_∞ , a najlepsze dla metryki L_1



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem średniokwadratowym w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. W tym przypadku widać niższe błędy dla metryki L1. W tym przypadku *moda* cechuje się największym błędem średniokwadratowym.



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru danych można zaobserwować że najlepiej radzi sobie funkcja *randomForest* z najmniejszymi wszystkimi błędami. Błąd ERR tej funkcji jest porównywalny z błędem 1-nn. Błędy funkcji *svm* są porównywalne z błędem knn dla $k = 3$ dla agregacji np. *moda*. Niestety użycie funkcji *polr* dało najgorsze wyniki porównywalne z knn dla dużego k , a MSE i MAD w tym przypadku było nawet gorsze.

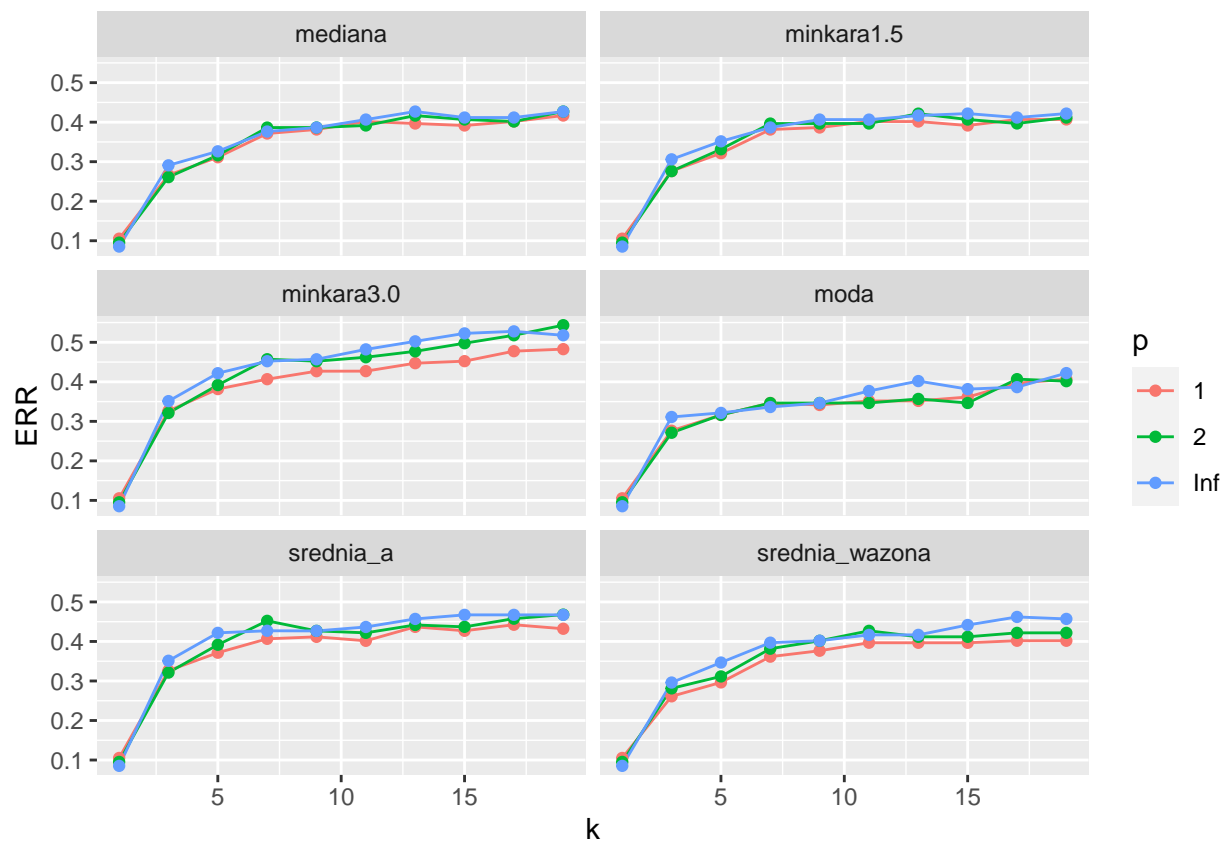
name	ERR	MAD	MSE
randomForest	0.1229249	0.2062988	0.4687495
polr	0.7343758	1.4976826	4.1688359
svm	0.4774169	0.7211923	1.4484908

Zbiór danych *machine_ord*

Poniżej znajdują się wykresy obrazujące następujące dane:

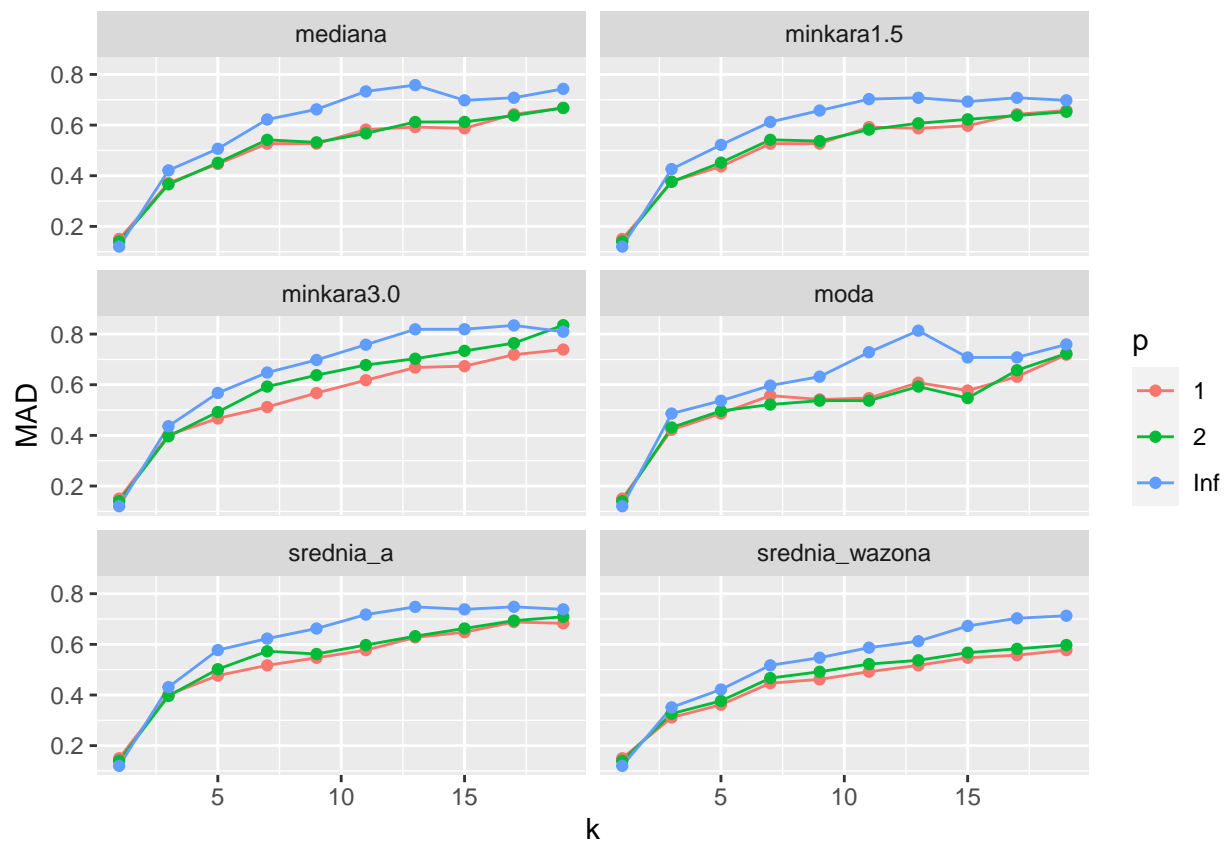
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że błąd ERR zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. Widać też że dla niektórych k zauważalnie zmniejsza się błąd dla metryki L1



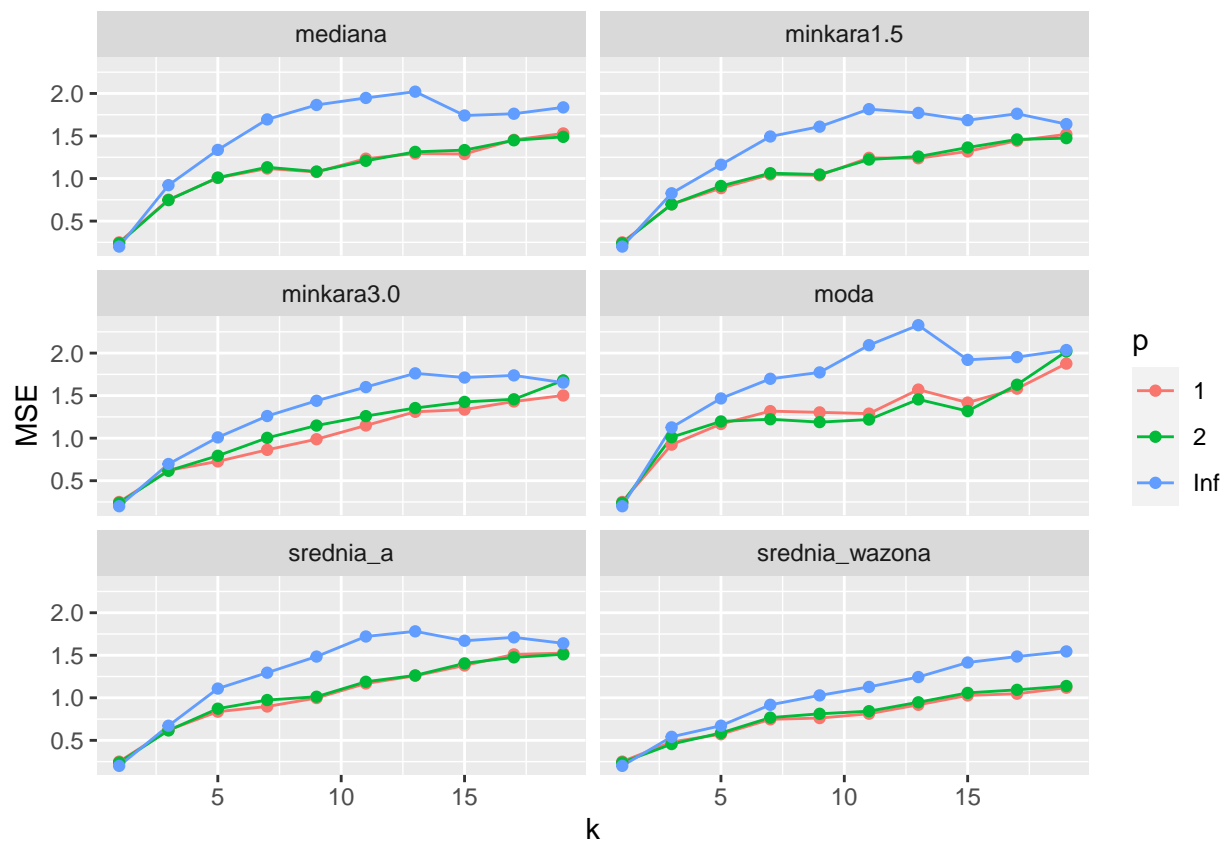
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem bezwzględnych w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . W tym przypadku widać że można zaobserwować niższe błędy dla metryki L1 i odrobinę wyższe dla metryki L2 i najwyższe dla L_∞ . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących, z wyjątkiem funkcji *moda*, która ma duże wahania w wartości. Ciekawe jest to że najmniejszym błędem średniokwadratowym i najłagodniejszym wzrostem w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Błąd wzrasta łagodniej niż w przypadku pozostałych zbiorów. W tym przypadku widać że można zaobserwować niższe błędy dla metryki L1, trochę większe dla L2 i największe dla L_∞ . W tym przypadku *moda* cechuje się największym błędem średniokwadratowym.



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru danych można zaobserwować że najlepiej radzi sobie funkcja *randomForest* z najmniejszymi wszystkimi błędami. Błąd ERR tej funkcji jest porównywalny z błędem 1-nn. Błędy funkcji *svm* są porównywalne z błędami knn dla $k = 3$. Niestety użycie funkcji *polr* było niemożliwe bez usuwania dodatkowych oprócz liniowo zależnych kolumn ze zbioru danych.

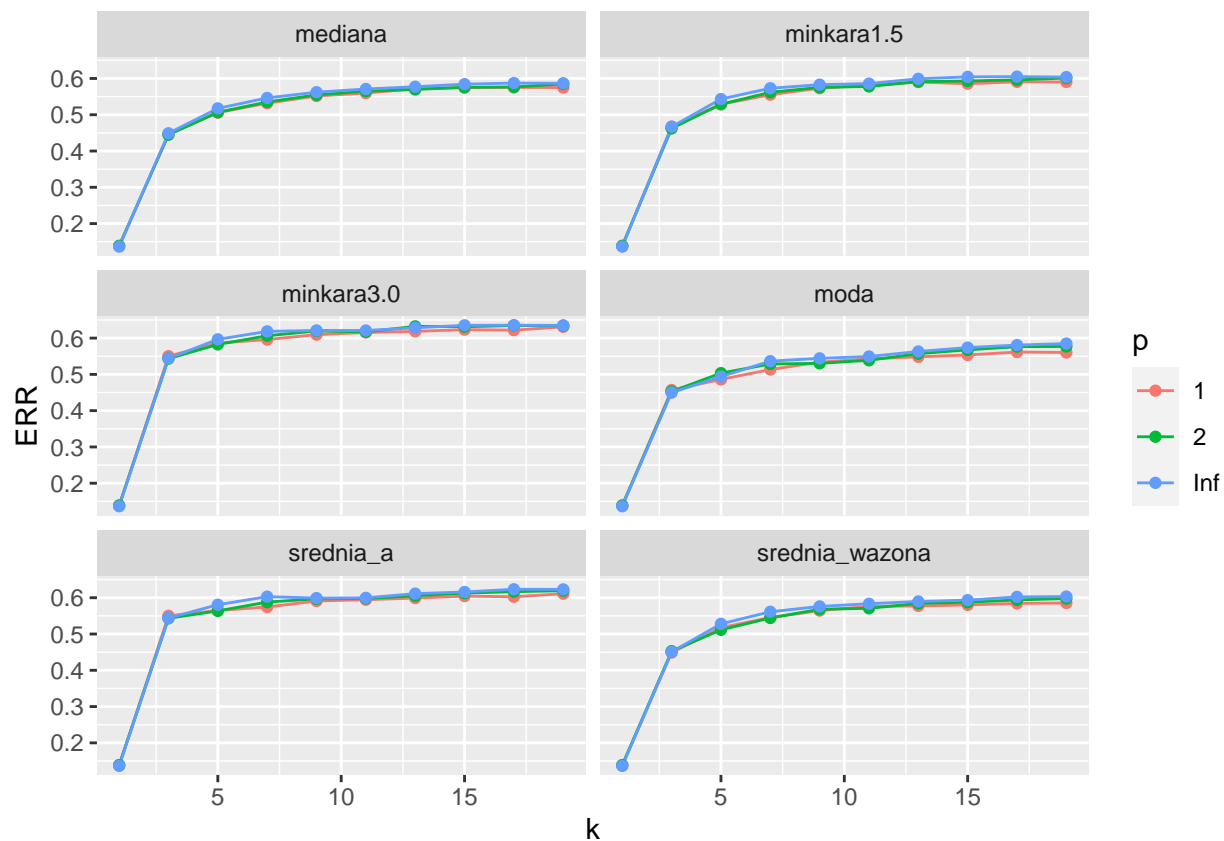
name	ERR	MAD	MSE
randomForest	0.09538462	0.1405128	0.2307692
svm	0.29602564	0.4315385	0.8328205

Zbiór danych *skill*

Poniżej znajdują się wykresy obrazujące następujące dane:

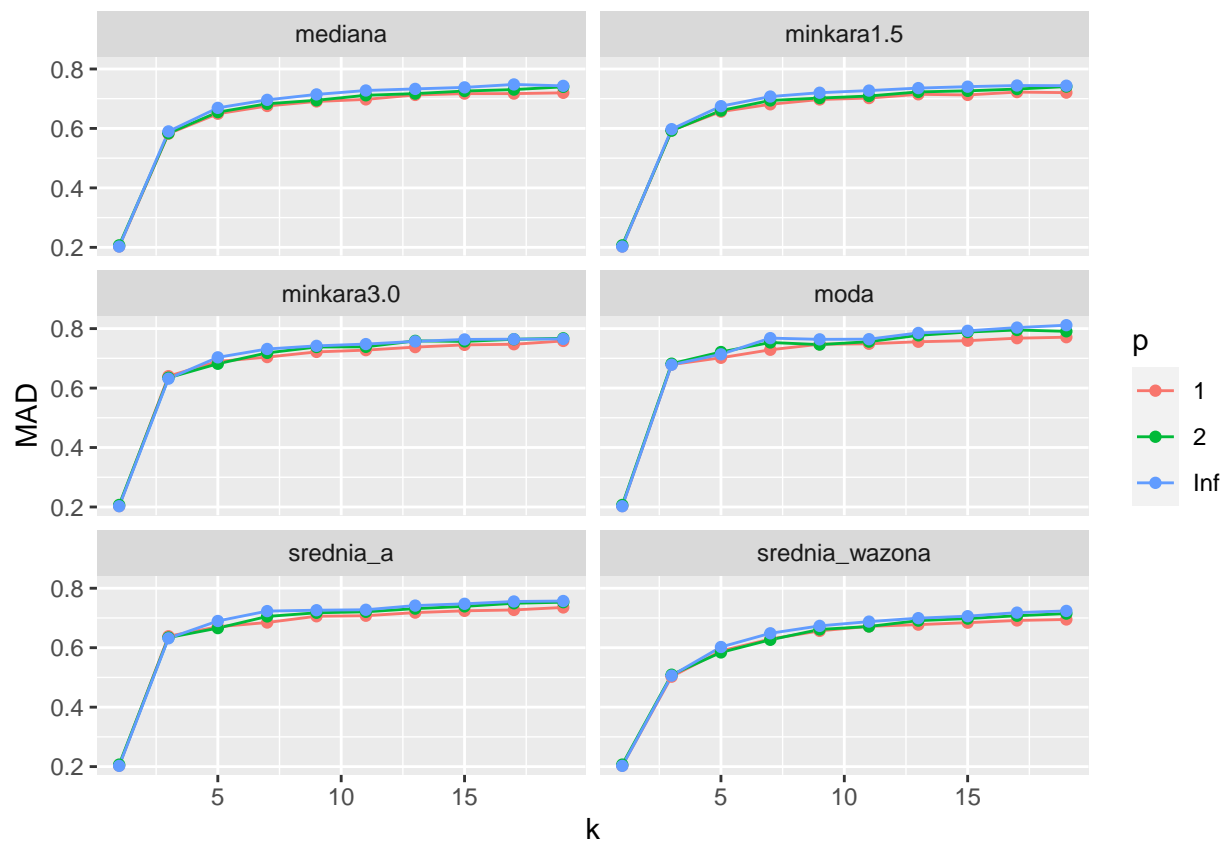
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że błąd ERR zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



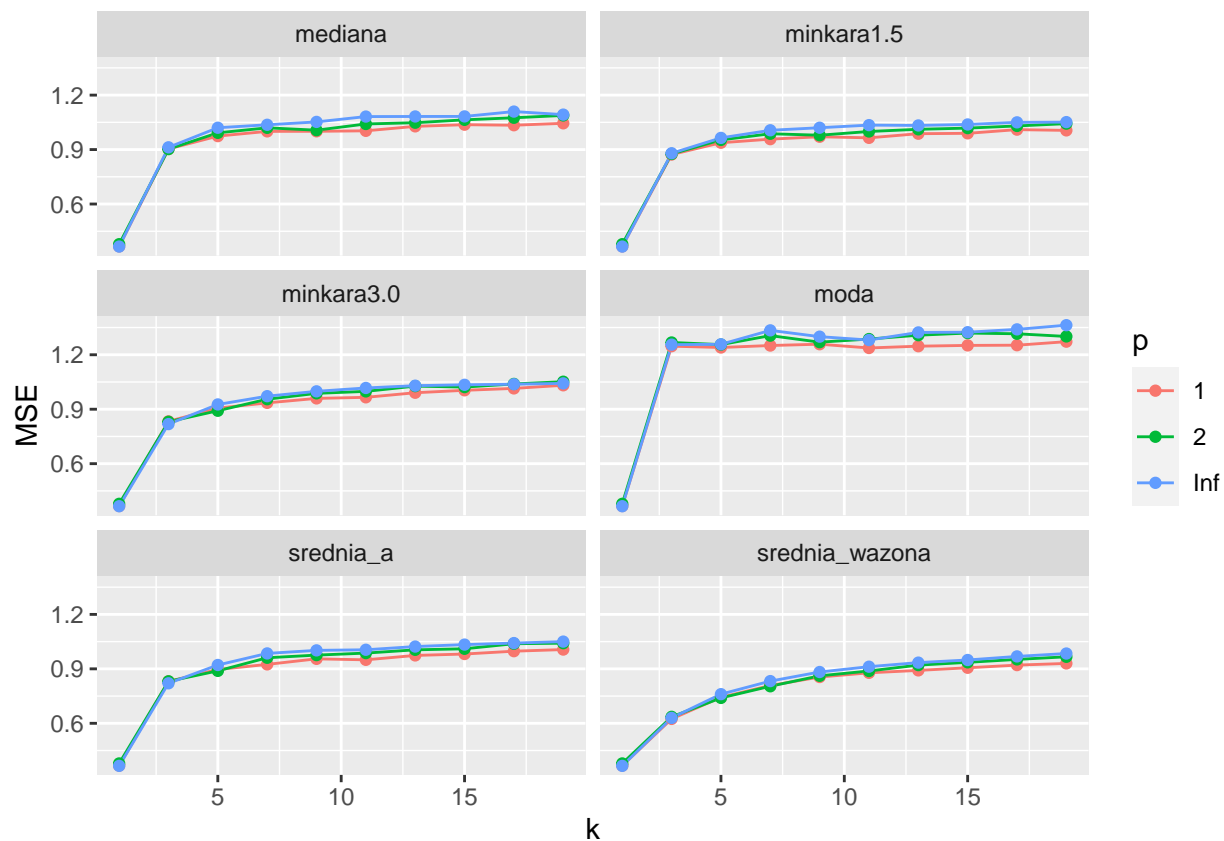
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem bezwzględnych w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem średniokwadratowym w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. W tym przypadku widać minimalnie niższe błędy dla metryki L1. W tym przypadku moda cechuje się największym błędem średniokwadratowym.



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru danych można zaobserwować że najlepiej radzi sobie funkcja *randomForest* z najmniejszymi wszystkimi błędami. Błąd ERR tej funkcji jest porównywalny z błędem 1-nn. Błędy funkcji *svm* są porównywalne z błędami knn dla $k = 3$. Niestety użycie funkcji *polr* było niemożliwe bez usuwania dodatkowych oprócz liniowo zależnych kolumn ze zbioru danych.

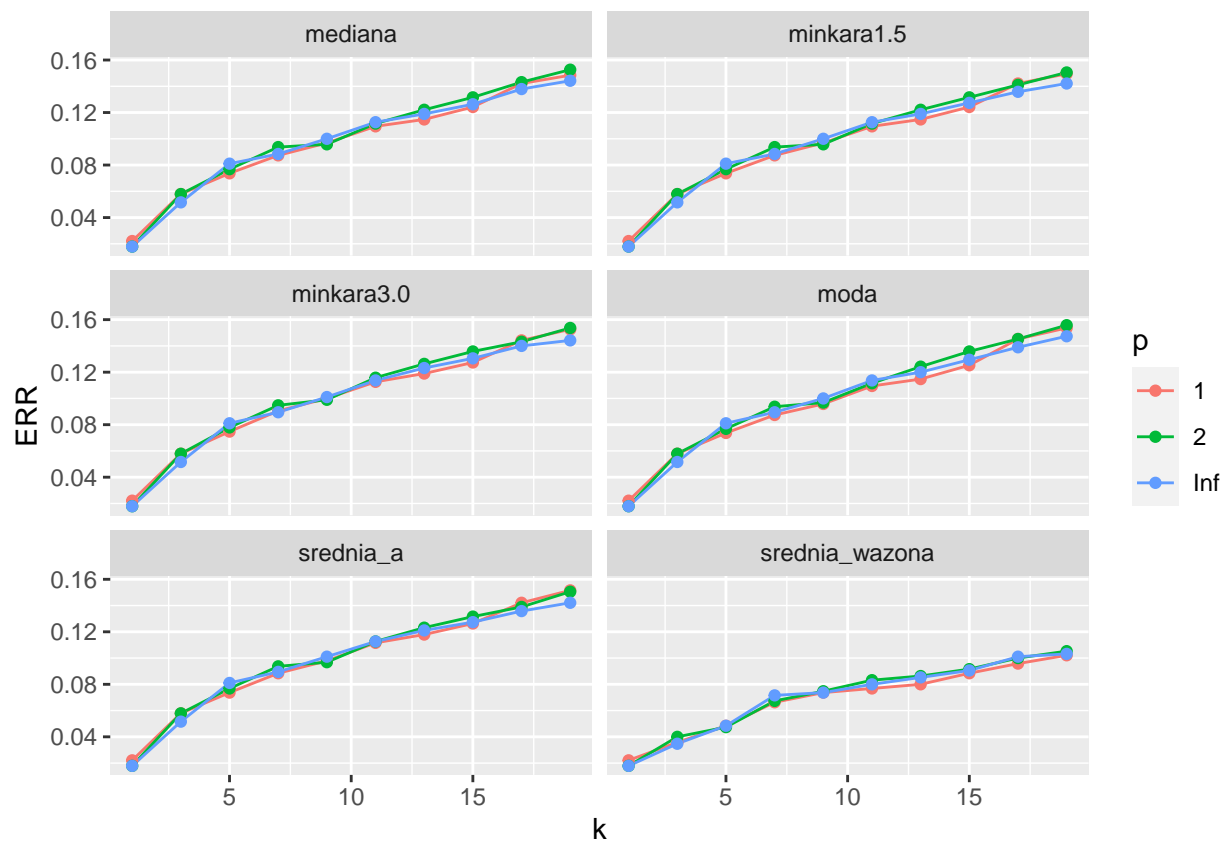
name	ERR	MAD	MSE
randomForest	0.1171754	0.1480411	0.2163638
svm	0.4941659	0.6488019	0.9970320

Zbiór danych *stock_ord*

Poniżej znajdują się wykresy obrazujące następujące dane:

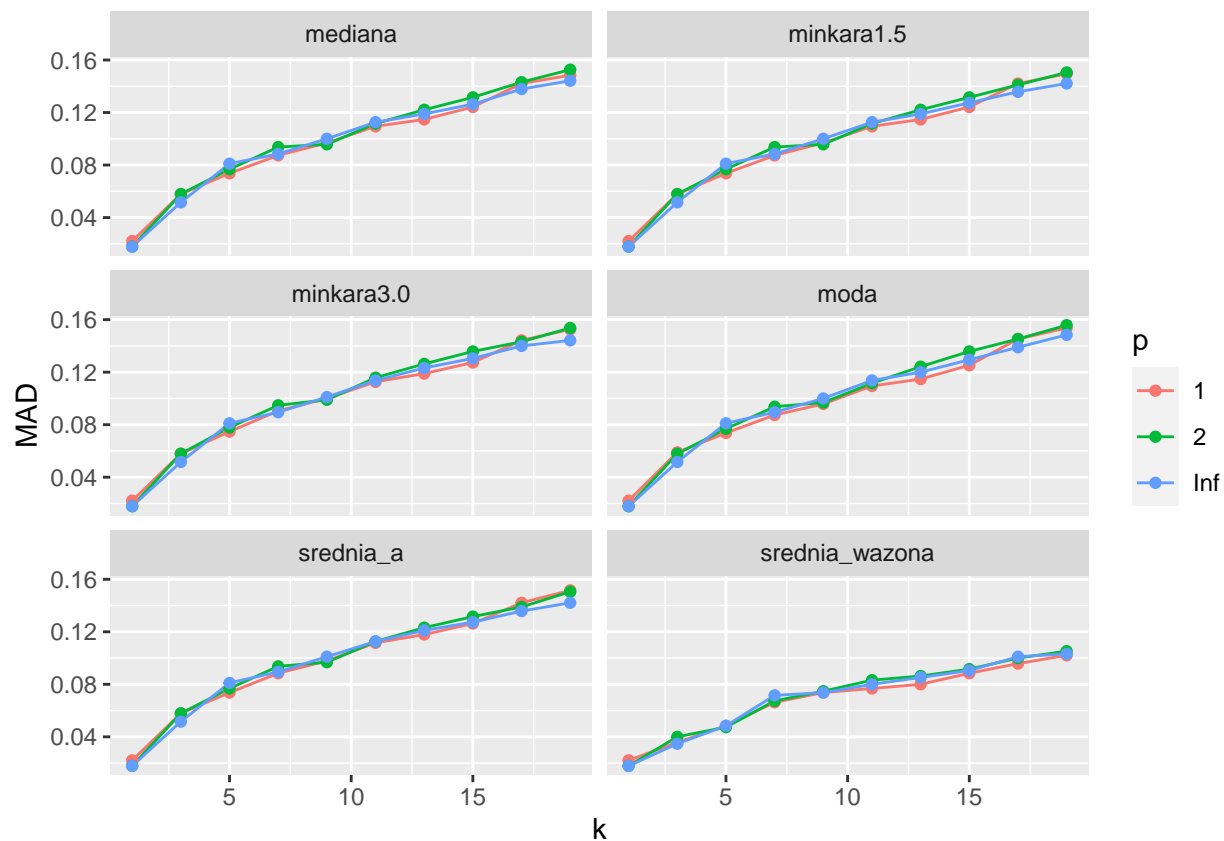
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że błąd ERR zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. Dla wszystkich funkcji wykresy te rosną łagodniej niż w innych zbiorach. Najlepsze wyniki daje funkcja *srednia_wazona*



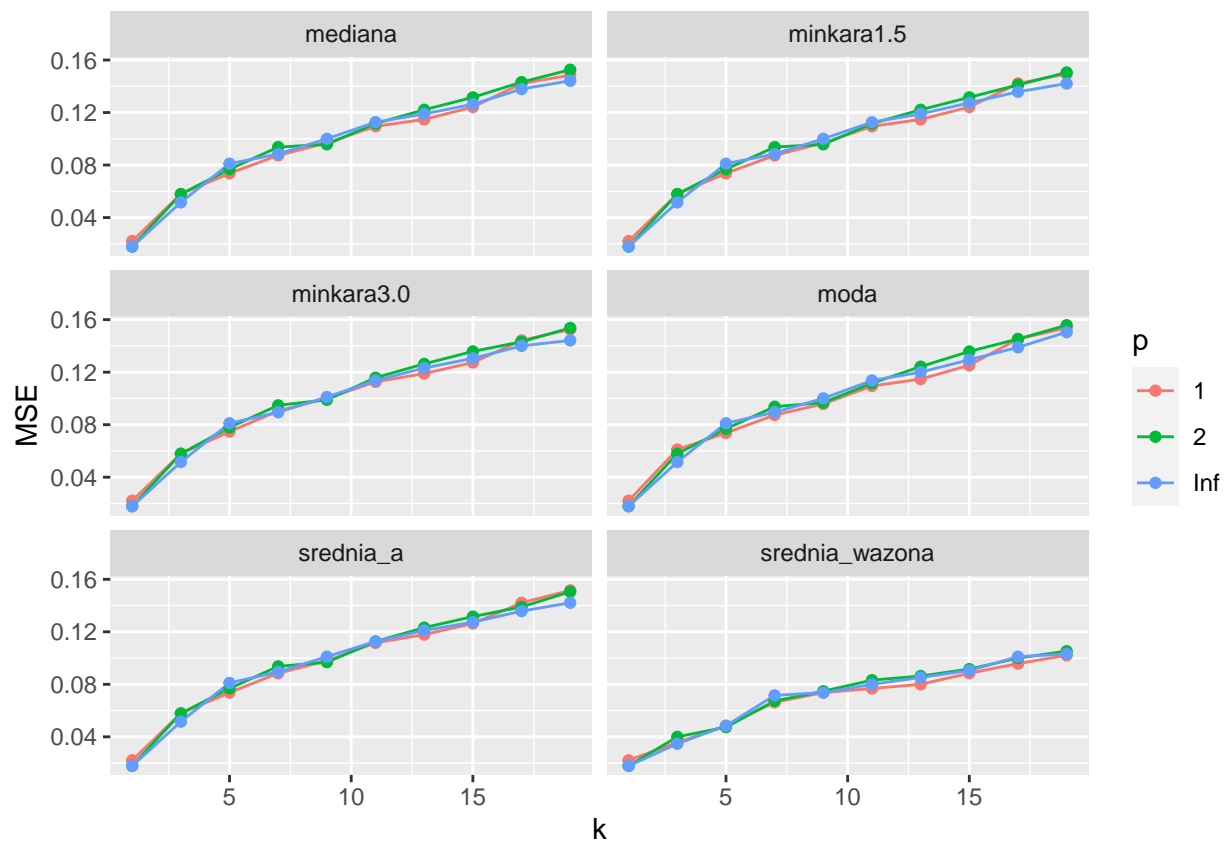
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem bezwzględnym w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . W tym przypadku widać że wszystkie metryki osiągają podobne wartości. Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem średniokwadratowym i najłagodniejszym wzrostem w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągniany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Błąd wzrasta łagodniej niż w przypadku pozostałych zbiorów. W tym przypadku widać że można zaobserwować niższe błędy dla metryki L1, trochę większe dla L2 i największe dla L_∞ . W tym przypadku moda cechuje się największym błędem średniokwadratowym.



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru danych można zaobserwować że najlepiej radzi sobie funkcja *randomForest* z najmniejszymi wszystkimi błędami. Błąd ERR tej funkcji jest porównywalny z błędem 1-nn. Błędy funkcji *svm* są porównywalne z błędami knn dla $k = 3$. Niestety użycie funkcji *polr* dała najgorsze wyniki, gorsze od funkcji knn.

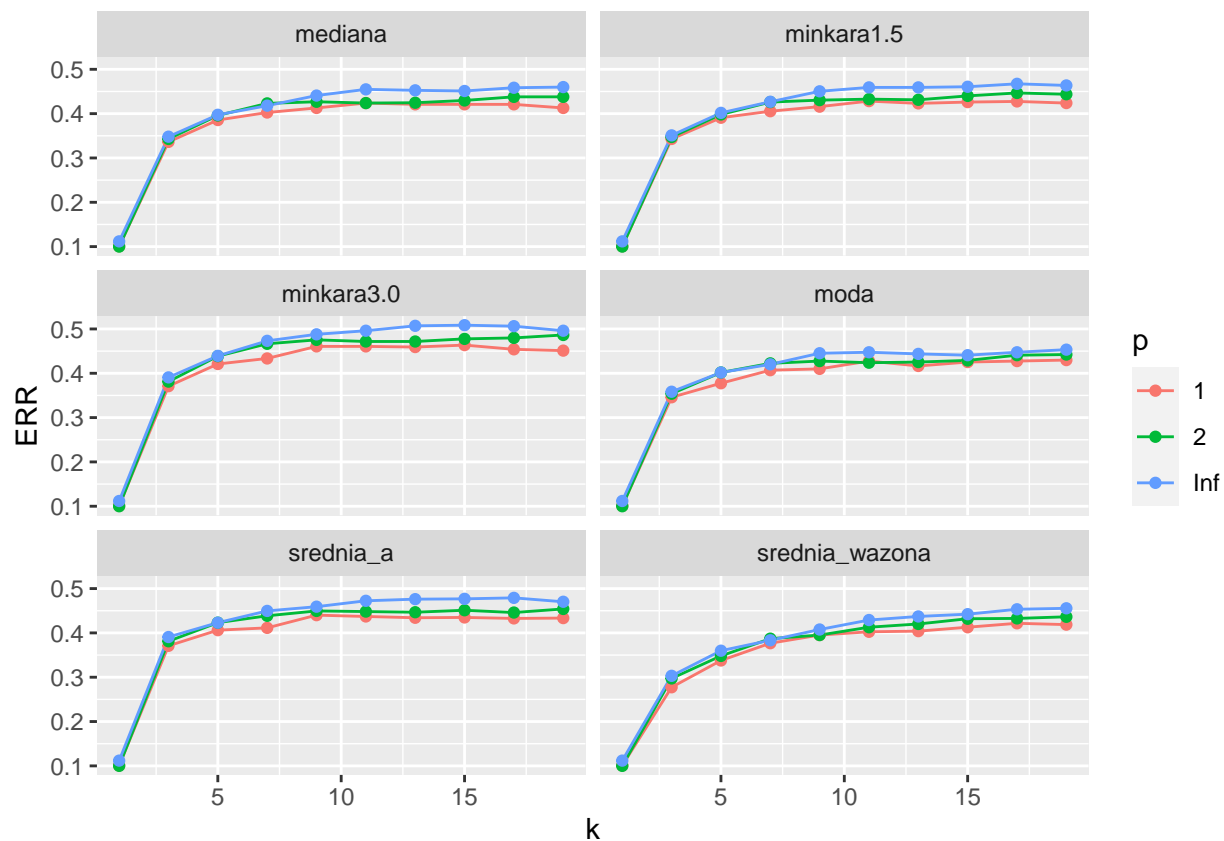
name	ERR	MAD	MSE
randomForest	0.01368421	0.01368421	0.01368421
polr	0.30000000	0.30736842	0.32210526
svm	0.16105263	0.16105263	0.16105263

Zbiór danych *winequality-red*

Poniżej znajdują się wykresy obrazujące następujące dane:

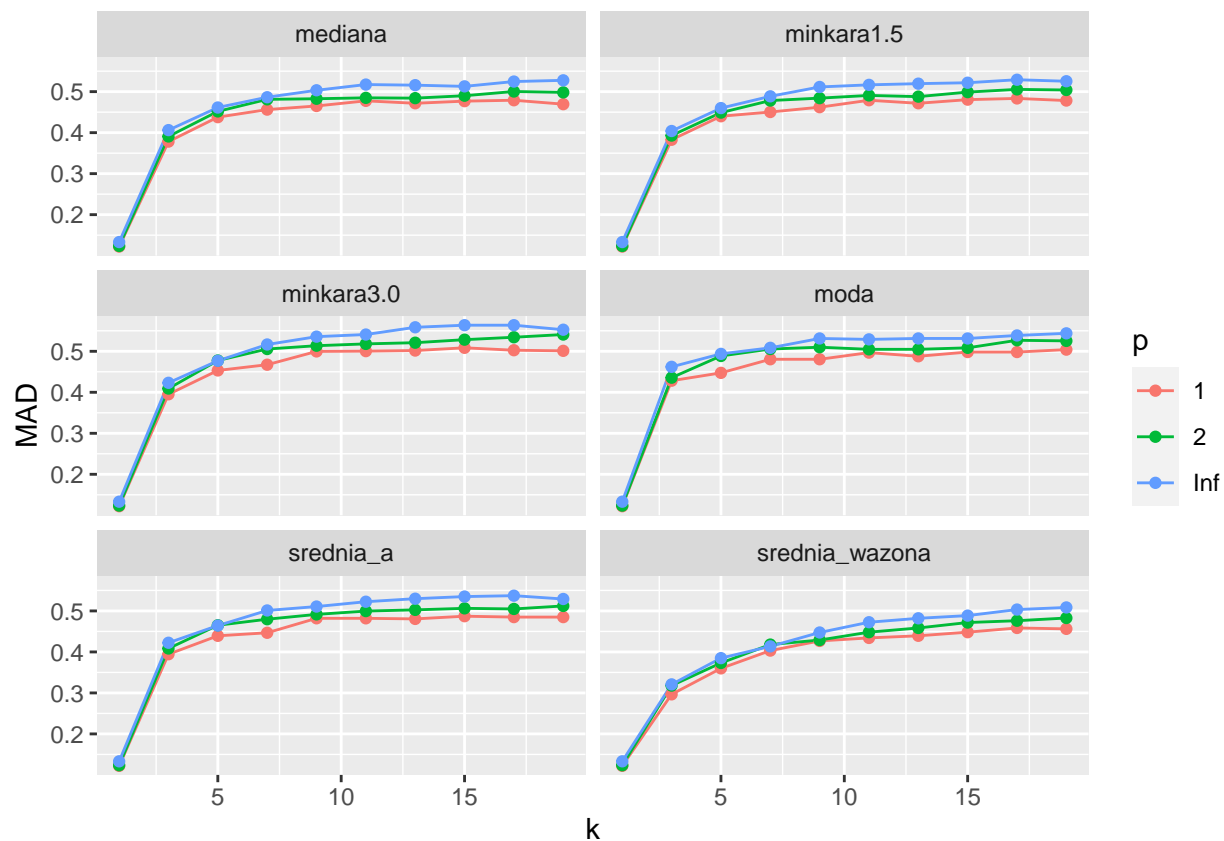
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że błąd ERR zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. Najlepsze wyniki dała funkcja *srednia_wazona*.



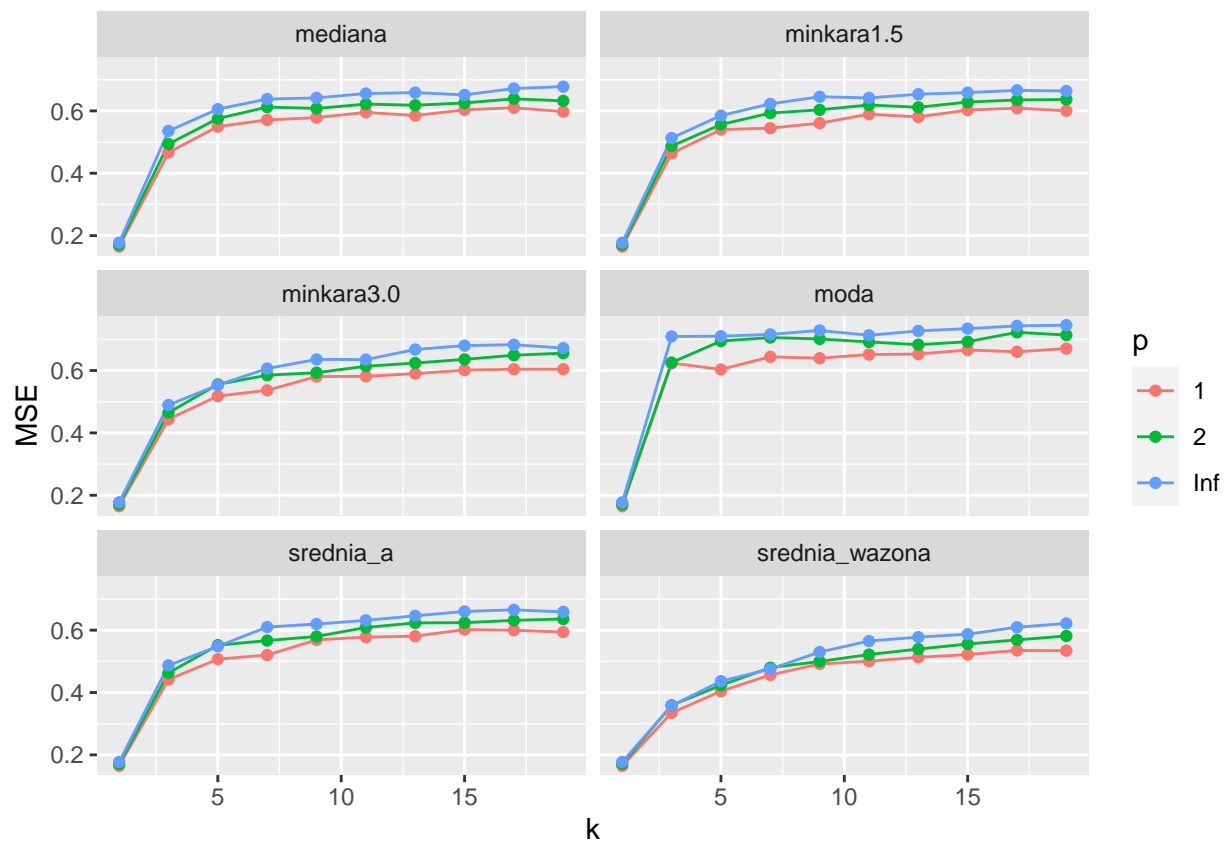
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem bezwzględnych w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. Najgorsze wyniki zostały uzyskane dla metryki L_∞ , a najlepsze dla metryki L_1



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem średniokwadratowym w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. W tym przypadku widać niższe błędy dla metryki L1. W tym przypadku *moda* cechuje się największym błędem średniokwadratowym.



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru danych można zaobserwować że najlepiej radzi sobie funkcja *randomForest* z najmniejszymi wszystkimi błędami. Błąd ERR tej funkcji jest porównywalny z błędem 1-nn. Błędy funkcji *svm* są porównywalne z błędem knn dla $k = 3$ dla agregacji *srednia_wazona*. Niestety użycie funkcji *polr* dało najgorsze wyniki porównywalne z knn dla dużego k .

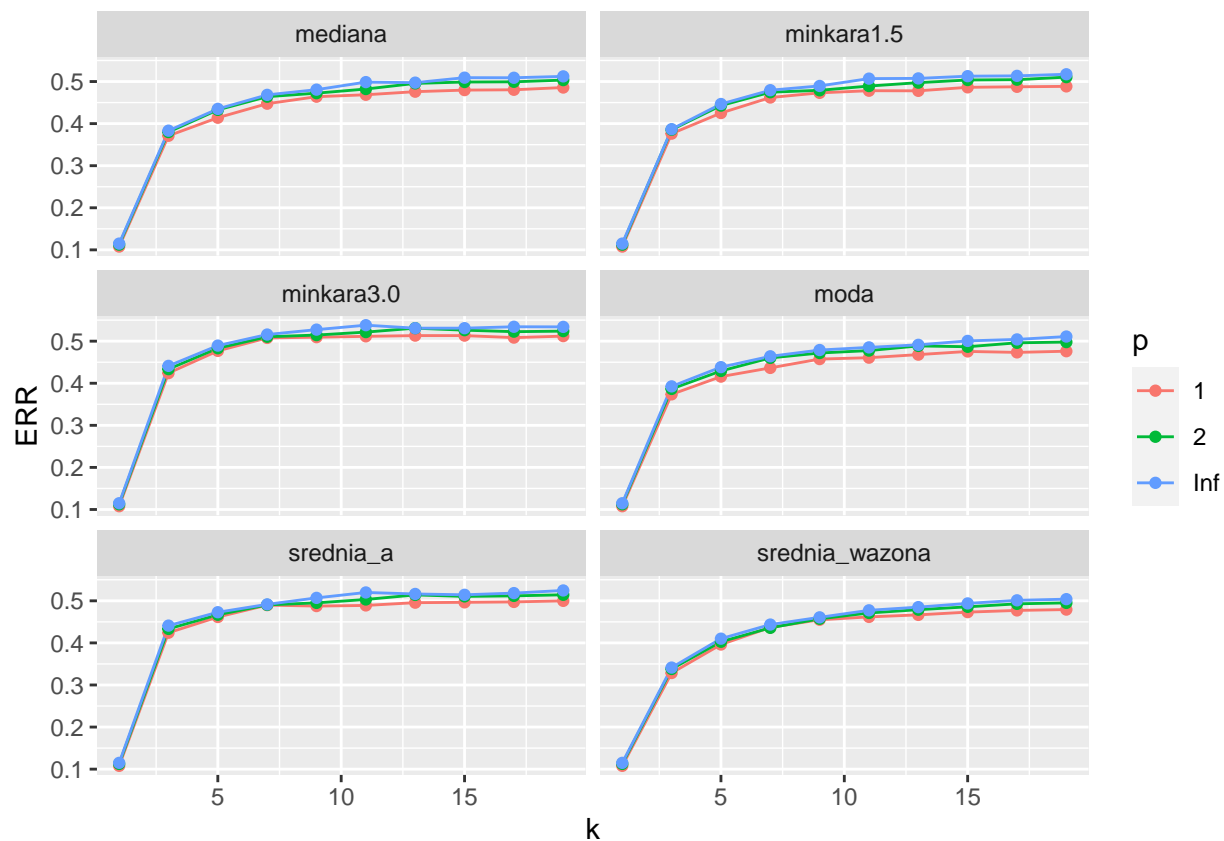
name	ERR	MAD	MSE
randomForest	0.07432711	0.08241535	0.09859182
polr	0.41943510	0.45475364	0.52833189
svm	0.35543195	0.39518667	0.48058932

Zbiór danych *winequality-white*

Poniżej znajdują się wykresy obrazujące następujące dane:

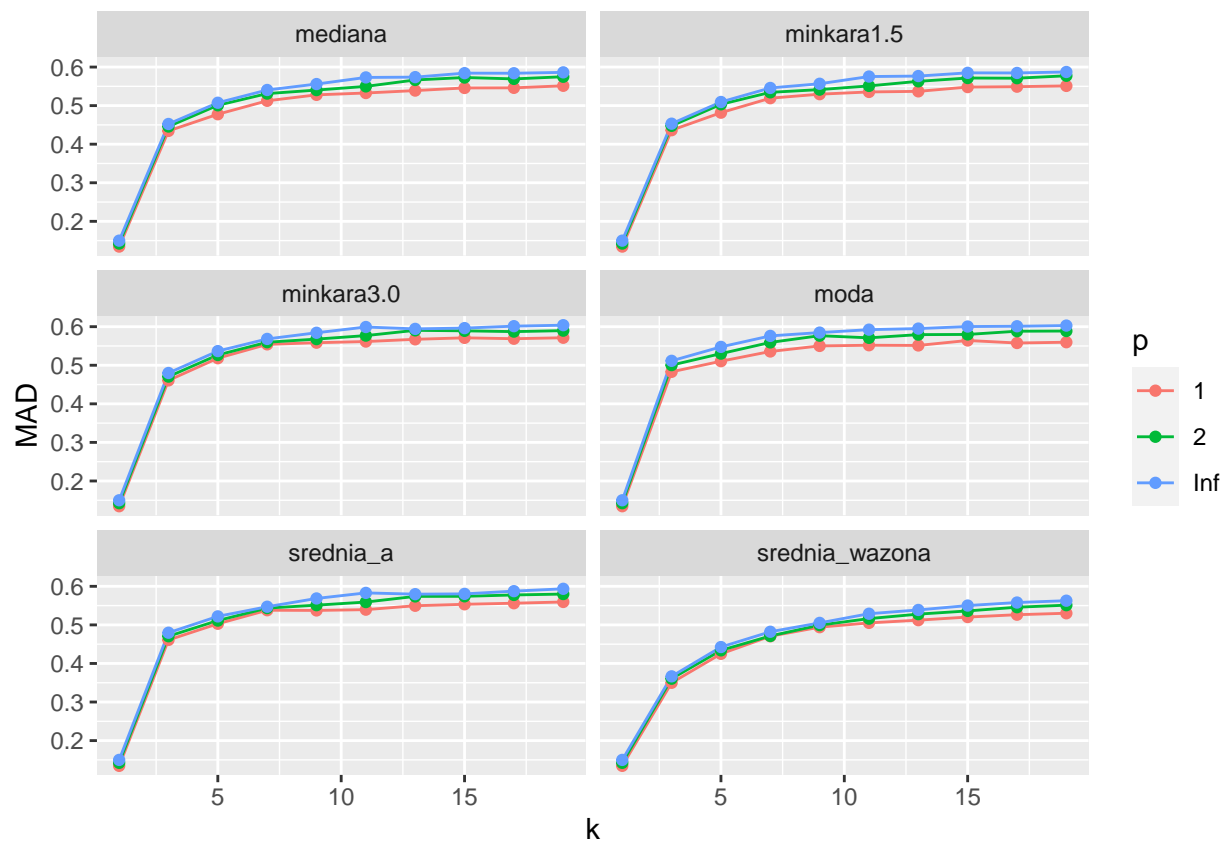
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że błąd ERR zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



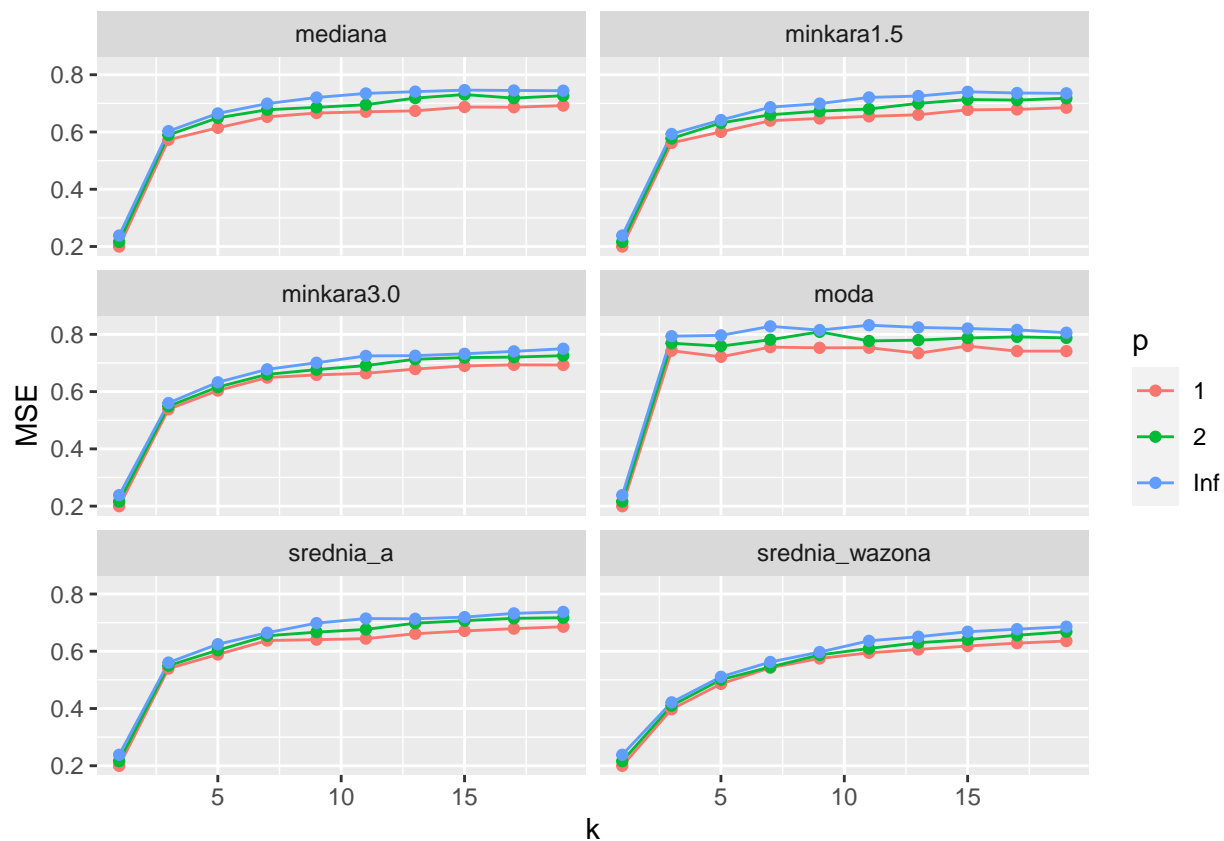
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem bezwzględnych w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. Najgorsze wyniki zostały uzyskane dla metryki L_∞ , a najlepsze dla metryki L_1



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest że najmniejszym błędem średniokwadratowym w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. W tym przypadku widać niższe błędy dla metryki L1. W tym przypadku *moda* cechuje się największym błędem średniokwadratowym.



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru danych można zaobserwować że najlepiej radzi sobie funkcja *randomForest* z najmniejszymi wszystkimi błędami. Błąd ERR tej funkcji jest porównywalny z błędem 1-nn. Błędy funkcji *svm* są porównywalne z błędem knn dla $k = 3$. Niestety użycie funkcji *polr* dało najgorsze wyniki porównywalne z knn dla dużego k .

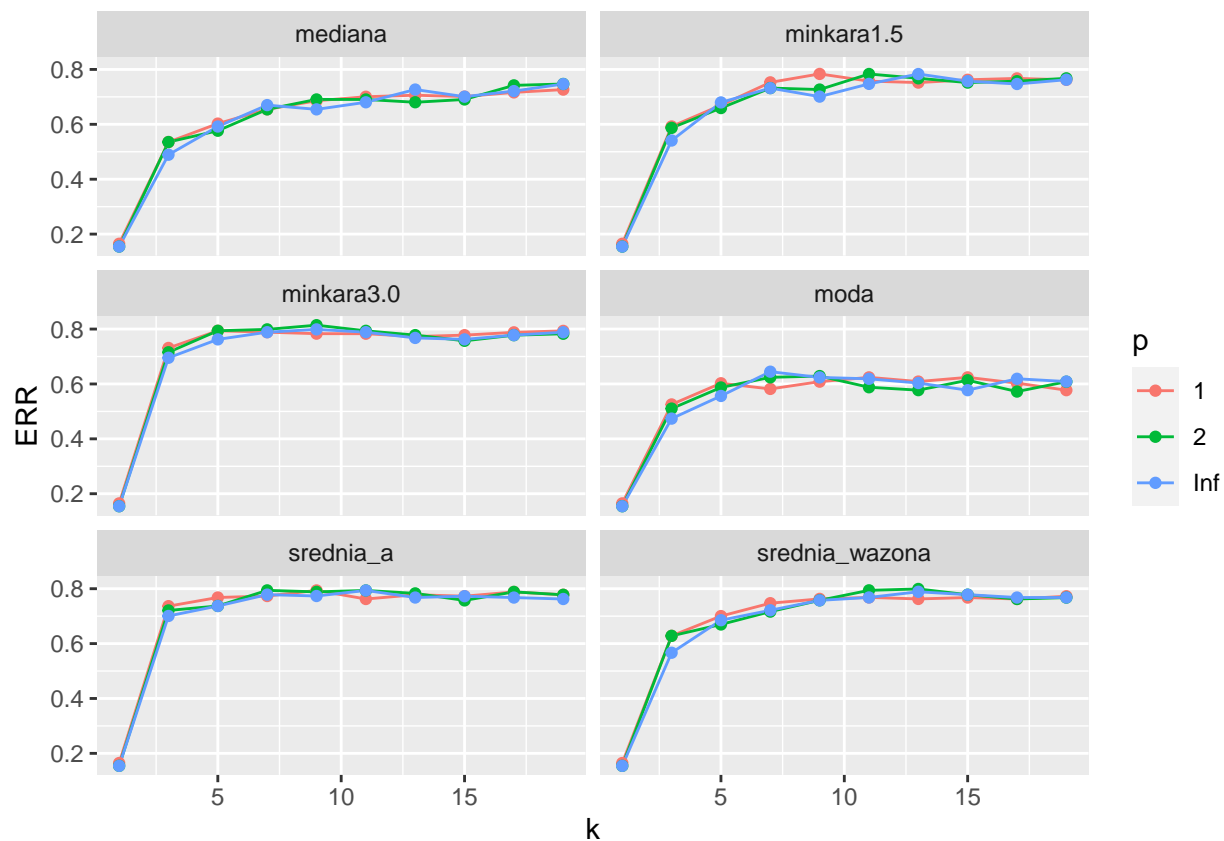
name	ERR	MAD	MSE
randomForest	0.08860866	0.09719133	0.1153668
polr	0.47109079	0.52865159	0.6558906
svm	0.40418784	0.45594087	0.5665138

Zbiór danych *wisconsin_breast_ord*

Poniżej znajdują się wykresy obrazujące następujące dane:

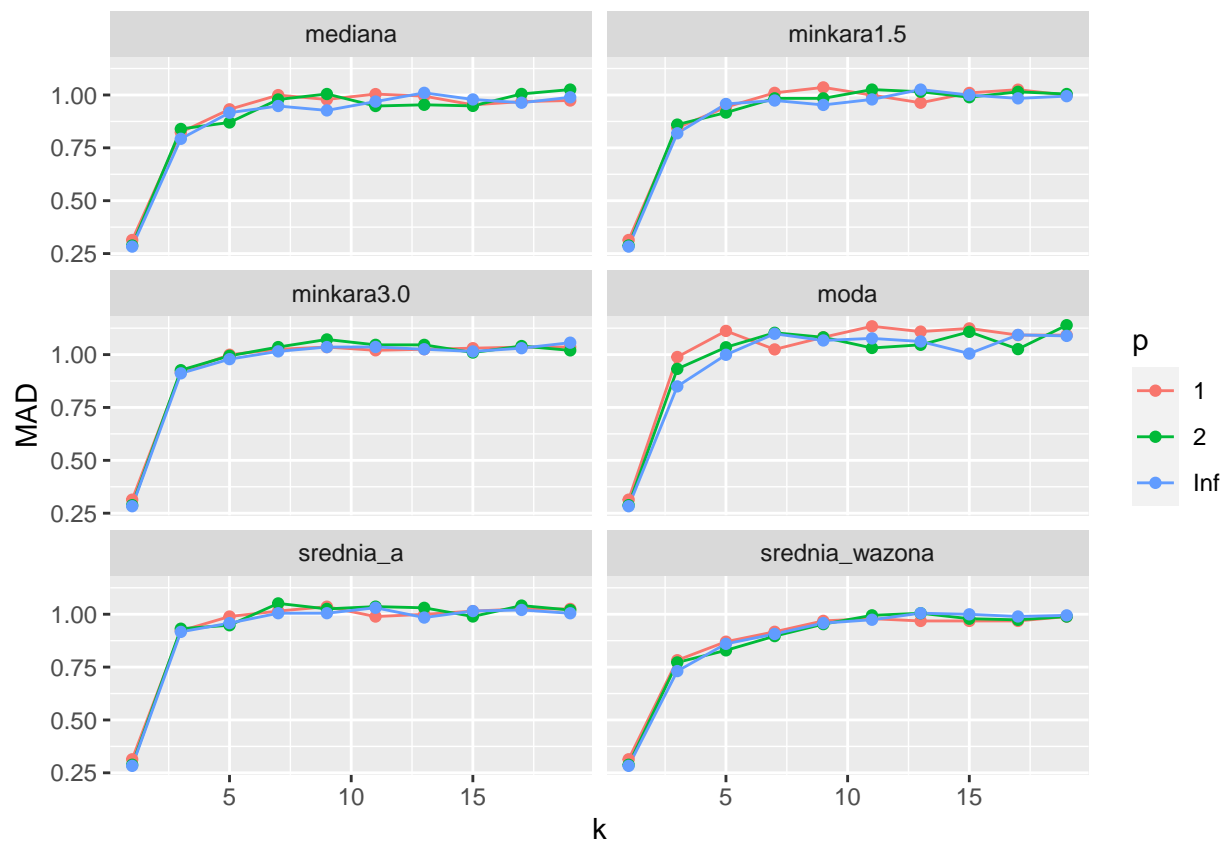
- **ERR** - proporcję błędnej klasyfikacji

Na poniższych wykresach widzimy że błąd ERR zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. Najlepsze wyniki daje funkcja *moda*.



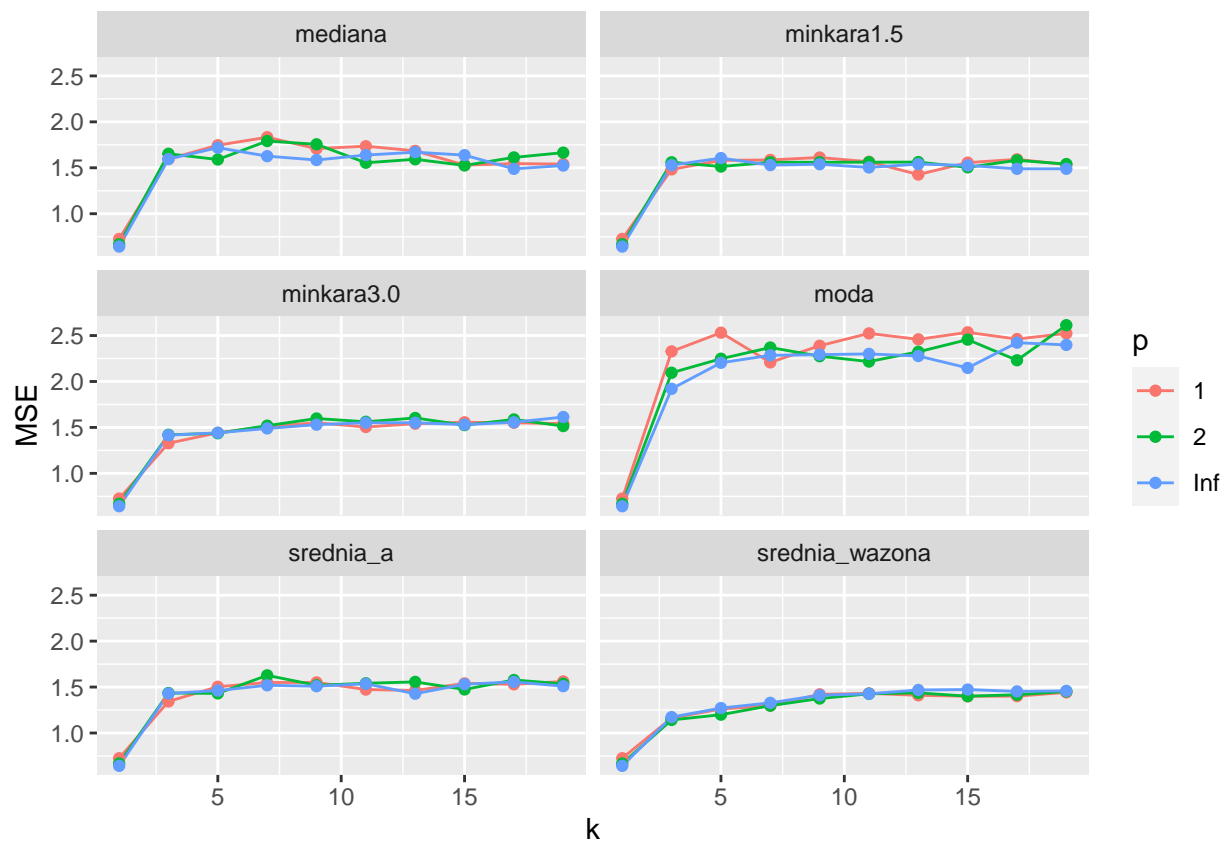
- **MAD** - błąd bezwzględny

Na poniższych wykresach widzimy że błąd MAD zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem bezwzględnych w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3.



- **MSE** - błąd średniokwadratowy

Na poniższych wykresach widzimy że błąd MSE zachowuje się w podobny sposób dla wszystkich funkcji agregujących. Ciekawe jest to że najmniejszym błędem średniokwadratowym w tym przypadku cechuje się funkcja *srednia_wazona*. Najmniejszy błąd osiągany jest przy $k = 1$, a następnie błąd wzrasta wraz z wzrostem k . Najgwałtowniejszy wzrost można zaobserwować dla k od 1 do 3. W tym przypadku *moda* cechuje się największym błędem średniokwadratowym.



Oraz tabela pokazująca te same błędy dla metod bibliotecznych

Dla tego zbioru danych można zaobserwować że najlepiej radzi sobie funkcja *randomForest* z najmniejszymi wszystkimi błędami. Błąd ERR tej funkcji jest porównywalny z błędem 1-nn. Błędy ERR pozostałych funkcji są porównywalne z błędami knn dla $k > 3$. Błędy funkcji *svm* są porównywalne z błędem knn dla $k = 3$. Niestety użycie funkcji *polr* dało najgorsze wyniki porównywalne z knn dla dużego k .

name	ERR	MAD	MSE
randomForest	0.1031039	0.1909582	0.4390013
polr	0.6033738	0.9699055	1.9202429
svm	0.4584345	0.9739541	2.5310391

Użycie funkcji *MASS:polr*

Z użyciem tej funkcji na zadanych zbiorach benchmarkowych pojawił się pewien problem związany prawdopodobnie z liniową zależnością pewnych kolumn w macierzy uczącej. Problem ten został rozwiązany z pomocą użycia funkcji *WeightIt::make_full_rank*, co prawda przez takie działanie usuwane są pewne cechy opisujące przypadki, ale niestety jest to konieczne do poprawnego działania funkcji *MASS:polr*.

Ogólny opis wyników

Po przeprowadzeniu wszystkich pomiarów, zobrazowaniu ich na różnorodnych tabelach można zaobserwować pewne łączące zależności między parametrami użycia funkcji knn do otrzymanych wyników. Oto one:

- Największa dokładność osiągnięta jest dla $k = 1$ w funkcji knn
Wynikać to może z faktu że nie zawsze inni najbliżsi sąsiedzi należą do dobrej grupy, kompletnie inna budowa zbiorów testowych i uczących również może zmienić tę zależność.
- W niewielu przypadkach użycie metryki L_∞ poprawia wyniki, w większości przypadków lepiej jest użyć metryki L_1
- W większości przypadków funkcja agregująca *moda* daje największe błędy średniokwadratowe.
- W większości przypadków funkcja agregująca *srednia_wazona* daje najmniejsze błędy średniokwadratowe.

Wynika to z tego że na wybranie etykiety najbardziej wpływa najbliższy sąsiad, który jak widać po pierwszej kropce jest najbardziej prawdopodobnym dobrym wyborem.

- Funkcja biblioteczna *randomForest* daje najlepsze wyniki bliskie pod względem ilości pomyłonych etykiet (ERR) do 1-nn oraz dużo lepsze przybliżenie tych które nie zostały dokładnie trafione błędy (MAD i MSE)
- Funkcja biblioteczna *e1071::svm* daje znacznie gorsze wyniki niż *randomForest*, ale przeważnie lepsze wyniki niż *MASS::polr*

Podsumowanie

Funkcja autorska **knn** będąca przedmiotem rozważań tego raportu cechuje się małą skutecznością dla dużej ilości wybranych najbliższych sąsiadów. O dziwo najlepsze efekty daje po prostu wybranie najbliższego sąsiada i użycie jego etykiety jako etykiety przedmiotu badanego. Istnieją metody biblioteczne takie jak *randomForest::randomForest* które cechują się podobnym współczynnikiem dobrze odgadniętych oraz błędami bezwzględnymi i średniokwadratowymi co knn dla $k = 1$.